CRYSTAL STRUCTURE, SUBSTRATE PROMISCUITY AND STRUCTURE-BASED ENGINEERING OF ARCHAEAL ISOPENTENYL PHOSPHATE KINASE

by

Mark Noel F. Mabanglo

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

University of Utah

May 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Mark Noel F. Mabanglo**

has been approved by the following supervisory committee members:

C. Dale Poulter	, Chair	**Dec. 8, 2011** Date Approved
Cynthia J. Burrows	, Member	**Dec. 8, 2011** Date Approved
Charles B. Grissom	, Member	**Dec. 8, 2011** Date Approved
Christopher P. Hill	, Member	**Dec. 8, 2011** Date Approved
Ryan E. Looper	, Member	**Dec. 8, 2011** Date Approved

and by **Henry S. White** , Chair of the Department of **Chemistry**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Archaea utilize the mevalonate (MVA) pathway for the biosynthesis of isopentenyl diphosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP), the two building blocks of all isoprenoids. The archaeal MVA pathway deviates from the classical eukaryotic MVA pathway in the two-step conversion of mevalonate phosphate to IPP by utilizing a putative phosphomevalonate decarboxylase and the recently discovered isopentenyl phosphate kinase (IPK). The latter catalyzes the ATP-dependent phosphoryl transfer reaction to isopentenyl phosphate (IP) to form IPP.

Chen and Poulter recently characterized IPKs from *Thermoplasma acidophilum* (THA) and *Methanothermobacter thermautotrophicus* (MTH), using its substrate IP and other small molecule isoprenoid monophosphates. In this dissertation, we report the first crystal structures of THA and MTH IPKs in complex with their substrates and products. The structures reveal key active site residues involved in substrate binding and catalysis. We also describe the promiscuity of IPK towards fosfomycin, an antibiotic produced in *Streptomyces* and the substrate of the fosfomycin resistance enzyme, FomA. The structure of FomA is highly homologous to that of IPK, indicating a possible evolutionary relationship between the two enzymes. Finally, we describe four THA IPK mutants with kinase activities toward the longer chain isoprenoids geranyl phosphate (GP) and farnesyl phosphate (FP), achieved by mutations located in the IP binding site.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

α	alpha
Å	Angstrom
AAK	amino acid kinase
AK	aspartokinase
AMPPNP	adenylyl-imidodiphosphate
APE	Alanine, proline, glutamate (motif)
Arc	Archaea
ASKHA	acetate and sugar kinases/Hsc70/actin
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
AU	absorbance units
β	beta
BC	biotin carboxylase
BCA	bicinchoninic acid
BSA	bovine serum albumin
°C	degrees Celsius
CCP4	Comprehensive Computing Suite for Protein Crystallography
CDP	cytidine diphosphate
CDP-ME	cytidylyldiphosphomethylerythritol

CDP-MEP	cytidylyldiphosphomethylerythritol phosphate
CHARMM	Chemistry at Harvard Molecular Mechanics
CK	carbamate kinase
CLD	chain length determination (domain)
cm	centimeter
cMEPP	cyclomethylerythritoldiphosphate
CoA	coenzyme A
CO_2	carbon dioxide
COOT	Crystallographic Object-Oriented Toolkit
C-spine	catalytic spine
CTP	cytidine triphosphate
δ	delta, used to denote chemical shift
dATP	deoxy-adenosine triphosphate
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DXP	deoxyxylulose phosphate
D_2O	deuterium oxide
ε	extinction coefficient
E	enzyme
EBI	European Bioinformatics Institute
EDTA	ethylenediamine tetraacetic acid
Fe	iron

FP	farnesyl phosphate
FPP	farnesyl diphosphate
FPPS	farnesyl diphosphate synthase
Fom	fosfomycin
FomA	fosfomycin kinase
FosA	fosfomycin resistance enzyme A
FosB	fosfomycin resistance enzyme B
FosX	fosfomycin resistance enzyme X
g	gram
G-loop	glycine-rich loop
GGPP	geranylgeranyl diphosphate
GGPPS	geranylgeranyl diphosphate synthase
GP	geranyl monophosphate
GPP	geranyl diphosphate
GROMACS	Groningen Machine for Chemical Simulations
GTP	guanosine triphosphate
G5K	glutamate-5-kinase
η	eta, used to denote 3_{10} helices
h	hour
HDMAPP	4-hydroxydimethylallyl diphosphate
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HGT	horizontal gene transfer
HMG CoA	hydroxymethylglutaryl CoA

HMGR	hydroxymethylglutaryl CoA reductase
HMGS	hydroxymethylglutaryl CoA synthase
HPPS	hexaprenyl diphosphate synthase
HRD	Histidine, arginine, aspartic acid (motif
IDI-1	isopentenyl diphosphate isomerase, type 1
IDI-2	isopentenyl diphosphate isomerase, type 2
IP	isopentenyl phosphate
IPK	isopentenyl phosphate kinase
IPP	isopentenyl diphosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
J	coupling constant
kJ	kilojoule
k_{cat}	catalytic constant
K_d	dissociation constant
K_{eq}	equilibrium constant
K_i	inhibition constant
K_M	Michaelis constant
λ	lambda, used to denote wavelength
L	liter
LB	Luria Bertani
LDH	lactate dehydrogenase
LDL	low density lipoprotein
LUCA	last universal common ancestor

MAD	multiwavelength anomalous dispersion
MALDI	matrix-assisted laser desorption/ionization
MEP	methylerythritol phosphate
mg	milligram
MIB	malonate/imidazole/borate (buffer)
min	minute
MJ	*Methanocaldococcus jannaschii*
MK	mevalonate kinase
mL	milliliter
mM	millimolar
mol	mole
MS	mass spectrometry
MTH	*Methanothermobacter thermautotrophicus*
MVA	mevalonate (pathway)
MWCO	molecular weight cutoff
NAD^+	nicotinamide adenine dinucleotide, oxidized form
NADH	nicotinamide adenine dinucleotide, reduced form
NADPH	nicotinamide adenine dinucleotide phosphate
NAG	N-acetylglutamate
NAGK	N-acetylglutamate kinase
Ni-NTA	nickel (II) nitriloacetic acid
nm	nanometer
NMR	nuclear magnetic resonance

NPT	constant particle number, pressure, temperature
ns	nanosecond
NVT	constant particle number, volume, temperature
OD	optical density
PCB	sodium propionate/sodium cacodylate/Bis-Tris propane (buffer)
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
pFTase	protein farnesyltransferase
pGGTase	protein geranylgeranyltransferase
PGK	phosphoglycerate kinase
P_i	phosphate
PISA	Protein interfaces, surfaces and assemblies
PK	pyruvate kinse
PMD	phosphomevalonate decarboxylase
PMK	phosphomevalonate kinase
PO_4	phosphate
PPMD	diphosphomevalonate decarboxylase
Pyr	pyruvate
R_f	retention factor
RMSD	root mean square deviation
rpm	rotations per minute

R-spine	regulatory spine
s	second
S	substrate
S	Sulfur
SAD	single wavelength anomalous dispersion
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SSRL	Stanford Synchrotron Radiation Lightsource
SW	*Streptomyces wedmorensis*
τ	tau, used to denote lag time
TEAP	triethylammonium phosphate
TEV	tobacco etch virus
THA	*Thermoplasma acidophilum*
TLC	thin layer chromatography
U	units of enzyme activity
μM	micromolar
UMPK	uridylate monophosphate kinase
UPLC	ultraperformance liquid chromatography
UTP	uridine monophosphate
V	catalytic rate
v/v	volume over volume
V_{max}	maximum catalytic rate
VO_3	vanadate

ACKNOWLEDGMENTS

This dissertation is a product of the efforts and guidance of many scientists and mentors. I would like to thank Dale, my advisor, for his unwavering kindness, generosity and understanding. I could not have chosen a better mentor for my graduate studies. I would also like to thank Christopher Hill and Heidi Schubert for the opportunity to work in a well-equipped protein crystallography laboratory that enabled us to solve protein structures. The rest of my committee, Cynthia Burrows, Charles Grissom, Ryan Looper and Gregory Voth, helped me develop into a student scientist with confidence in my abilities and potential.

I appreciate the help and friendship of past and present members of the Poulter and Hill laboratories. Along the way, they have helped me accomplish my goals in my graduate studies. I also extend my appreciation to my fellow Filipino students at the University of Utah for their support and friendship.

This dissertation is dedicated to my parents and siblings, to my best friend, Eric, and most importantly, to God.

CHAPTER 1

INTRODUCTION

Background and Significance

The terpenoids, also known as isoprenoids, are a diverse family of naturally-occuring organic compounds derived from five-carbon isoprene units assembled and modified in thousands of ways to form mostly polycyclic structures with different carbon skeletons and functional groups. More than 60,000 terpenoids are known to date, acting as hormones (gibberellins, progesterone), structural components of cell membranes (cholesterol, ergosterol, hopanoids), electron transporters (haem A, ubiquinone), visual pigments (retinal), polysaccharide carriers for glycoprotein synthesis (dolichols), and protein modifiers for cellular localization (farnesylation and geranylgeranylation of Ras) (Figure 1.1) *(1)*. In insects, two forms of the terpenoid farnesene act as *semiochemicals*: α-farnesene is an alarm pheromone in termites and food attractant for the coddling moth *(2,3)*, while β-farnesene is an alarm pheromone in aphids that is released upon death *(4)*. The plant genus *Gardenia* also produces α-farnesene as the chief chemical contributor to its scent *(5)*. Other plant terpenoids act as photosynthetic pigments (chlorophyll), antimicrobial agents (phytoalexins) and as principal components of essential oils such as geraniol (rose perfumes), menthol (peppermint oil), citral (lemongrass oil), limonene

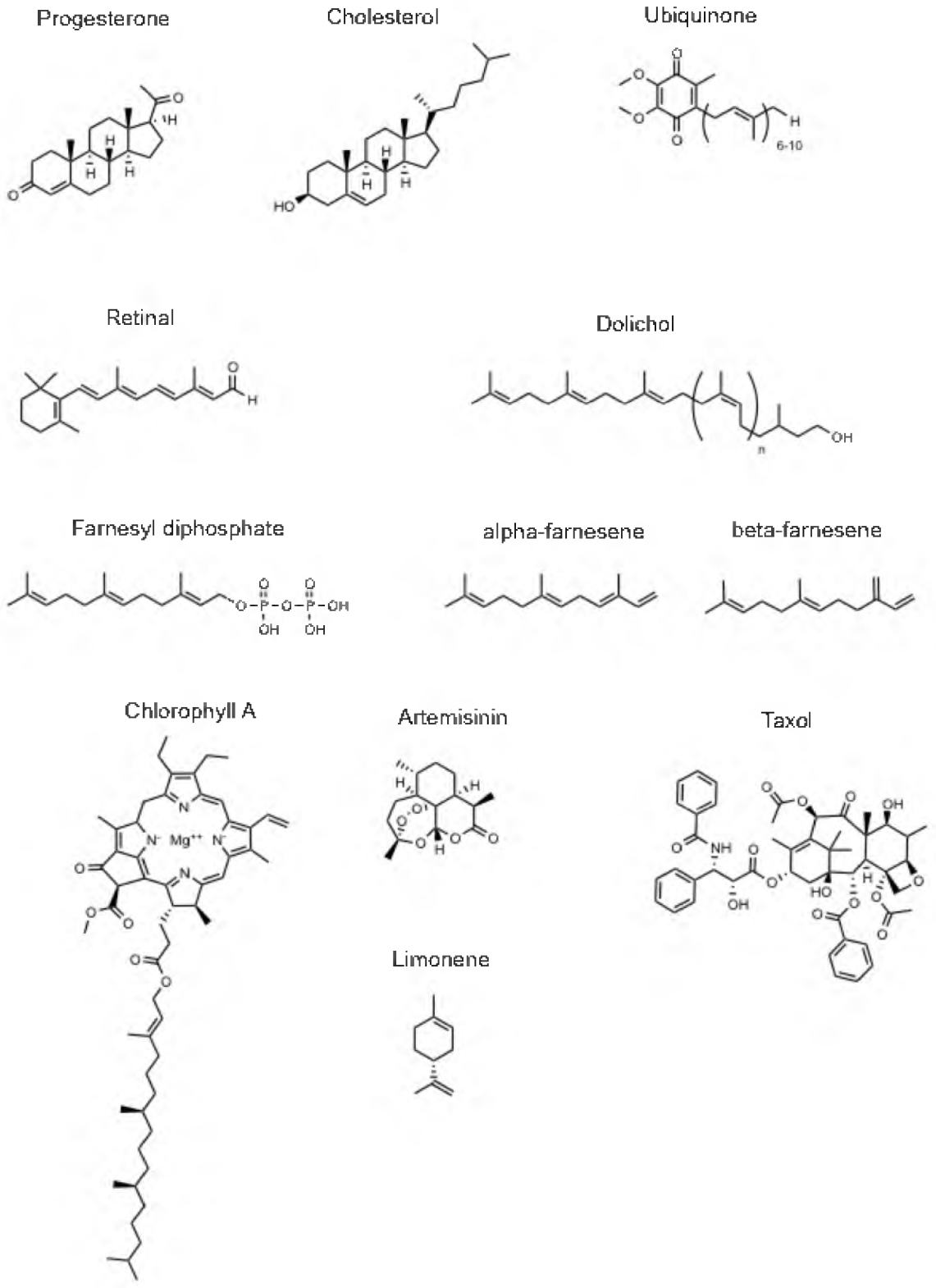


Figure 1.1 Chemical structures of representative examples of isoprenoids.

(lemon and orange oils) and camphor (incense). Some terpenoids are used in the treatment of cancer as well as protozoan, fungal and bacterial infections. For instance, artemisinin derivatives possess the most rapid action against falciparum and vivax malaria *(6,7)*, while taxol, first isolated from the Pacific yew tree *Taxus brevifolia*, is a mitotic inhibitor *(8)* used in the treatment of lung, ovarian, breast, head and neck cancer, and advanced forms of Kaposi's sarcoma *(9)*. The industrial manufacture of rubber and gutta-percha, perfumes, agrochemicals and food flavorings also depend on terpenoids as starting materials. Almost all organisms depend on terpenoids for surivival except for some endosymbiotic bacteria with very small genomes, as well as the archaeon *Nanoarchaeum equitans* that lives in an obligatory association with the hyperthermophilic crenarchaeote *Ignicoccus hospitalis* to obtain its lipids from the latter *(10)*.

Despite a vast array of chemical structures and functions, all terpenoid carbon backbones are derived from the five-carbon isoprene units isopentenyl diphosphate (IPP) and dimethylallyl disphosphate (DMAPP). The latter serves as the allylic primer for the head-to-tail condensation of IPP catalyzed by prenyltransferases, and subsequent additions of IPP produce the longer-chain isoprenoids geranyl diphosphate (GPP, C_{10}), farnesyl diphosphate (FPP, C_{15}), geranylgeranyl diphosphate (GGPP, C_{20}) and longer chain polyprenyl diphosphates *(11,12)*. These compounds lie at major branch points in the isoprenoid biosynthetic pathway from which thousands of different polycyclic and acyclic terpenoids are generated *(13,14)*.

Two independent and nonhomologous pathways for the biosynthesis of terpenoids are known: the mevalonate (MVA) pathway (Figure 1.2) and the 1-deoxy-D-xylulose-5-

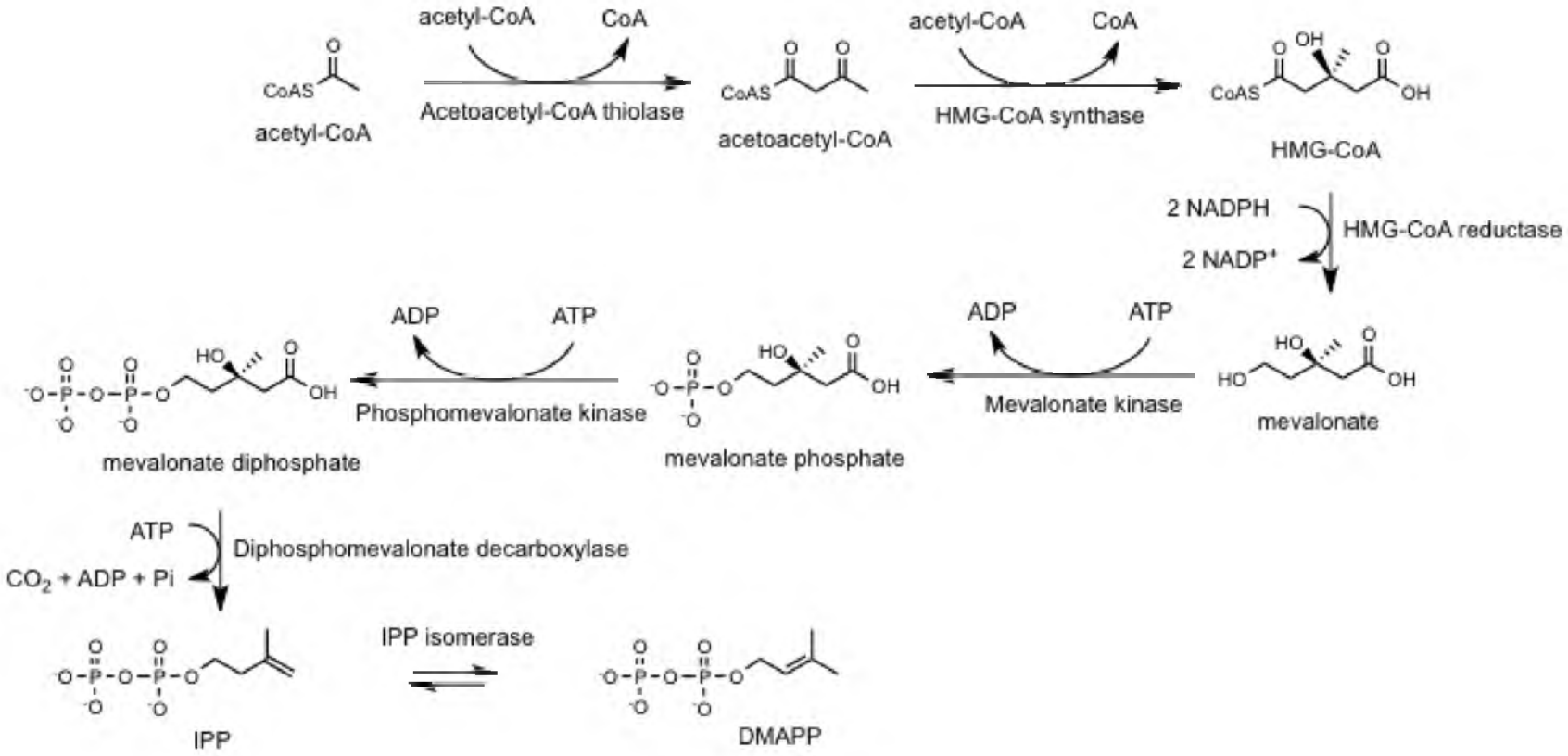


Figure 1.2 The classical MVA pathway for isoprenoid biosynthesis.

phosphate (DXP) pathway (Figure 1.3) *(15)*. The MVA pathway is found in most eukaryotes (all mammals), fungi, archaea, some eubacteria, the mitochondria and cytosol of plants and *Leishmania* and *Trypanosoma*. The DXP pathway operates in plant chloroplasts, algae, cyanobacteria, eubacteria, and apicomplexan parasites.

The MVA Pathway

The MVA pathway was first elucidated in the 1950s by Bloch, Cornforth and Lynen for the biosynthesis of cholesterol in yeast and animals *(15,16)*. As summarized in Figure 1.2, the pathway begins with the condensation of two molecules of acetyl CoA in a reaction catalyzed by acetoacetyl CoA thiolase. A third acetyl CoA is then condensed with acetoacetyl CoA by HMG CoA synthase to form 3-hydroxy-3-methylglutaryl-CoA (HMG CoA). The NADPH-dependent HMG CoA reductase then converts the CoA derivative to mevalonate, the rate-limiting step in the MVA pathway in humans. In ATP-dependent mechanisms, mevalonate then undergoes two successive phosphorylation steps catalyzed by mevalonate kinase (MK) and phosphomevalonate kinase (PMK). The diphosphate is subsequently decarboxylated by diphosphomevalonate decarboxylase (PPMD) to form IPP. To complete the MVA pathway, some of the IPP is converted to DMAPP by IPP isomerase.

Mevalonate, the product of HMG CoA reductase, is a key intermediate in the MVA pathway. It is used in the manufacture of cholesterol in the liver and intestines, and high cholesterol levels in the blood can damage arteries leading to cardiovascular diseases. To meet the physiological need for nonsterol isoprenoids without leading to cholesterol overaccumulation, several regulatory mechanisms exist, including the

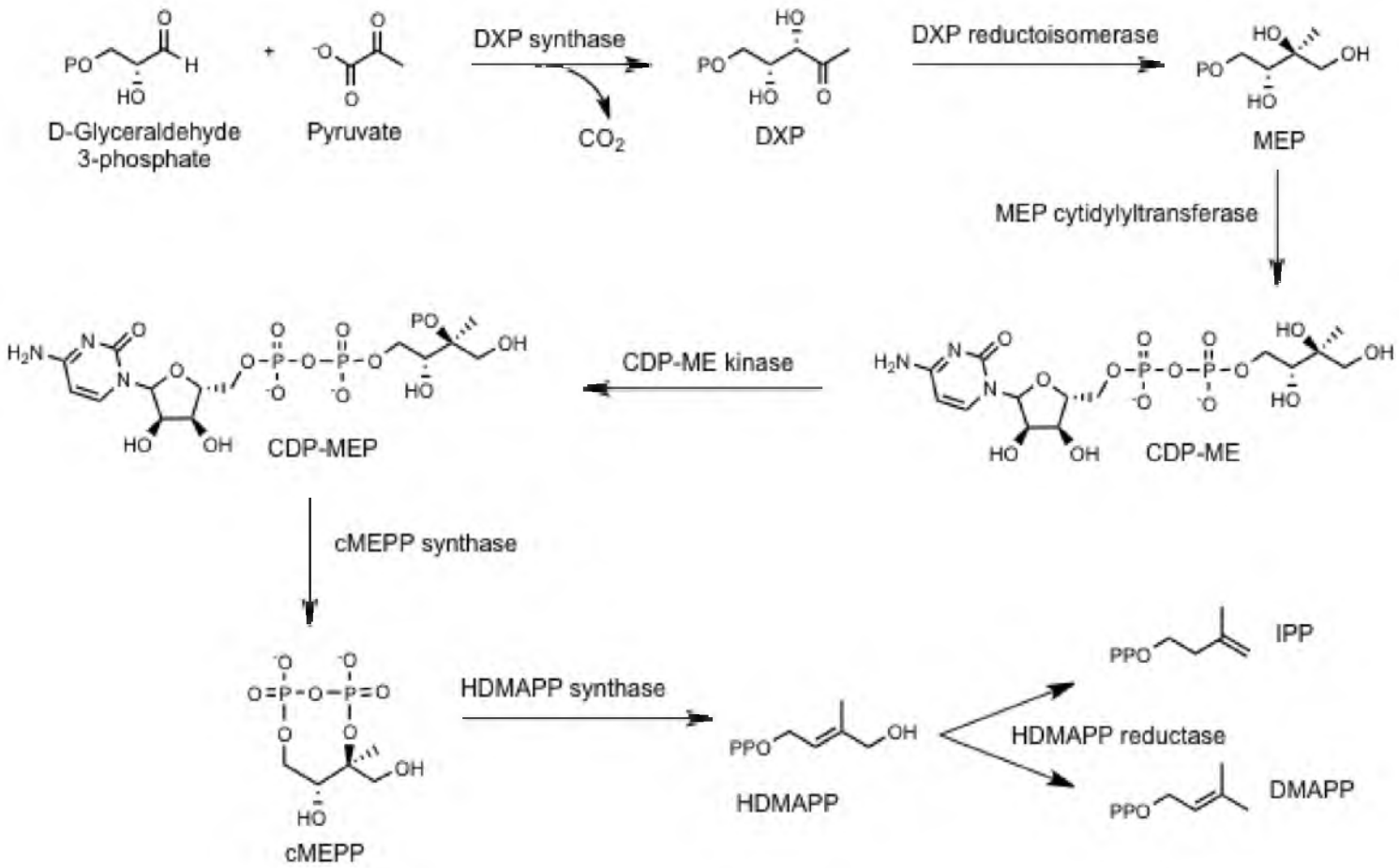


Figure 1.3. The DXP or MEP pathway for isoprenoid biosynthesis.

feedback regulation of HMG CoA synthase, HMG CoA reductase, and the LDL receptor *(11)*. In the absence of LDL, HMG CoA synthase and HMG CoA reductase have high activities, producing sufficient mevalonate for the synthesis of cholesterol and nonsterol isoprenoids (17). When LDL levels are high, the activities of the two enzymes are reduced by up to 90% and only small amounts of mevalonate are produced for nonsterol end products. When excess mevalonate is supplied externally together with LDL, residual HMG CoA reductase activity is abolished, terminating mevalonate production. The rise in sterol levels or decline in cholesterol demand results in the repression of the LDL receptor gene, preventing cholesterol overaccumulation *(18)*. Other mechanisms can also regulate mevalonate biosynthesis, including sterol-mediated transcription regulation *(19-25)* and posttranscriptional regulation of HMG CoA reductase *(26-29)*. Phosphorylation of HMG CoA reductase also regulates the enzyme's activity and can influence activities of other enzymes downstream *(29-31)*.

Cells not only regulate mevalonate synthesis but also control mevalonate disposition. Nonsterol enzymes generally have higher affinities than sterol enzymes for mevalonate-derived substrates *(17)*. At limiting concentrations of mevalonate, the synthesis of nonsterol isoprenoids is preferred. The committed step in sterol biosynthesis catalyzed by squalene synthetase is also suppressed upon prolonged incubation with sterols, thereby decreasing the incorporation of mevalonate into sterols.

With the central role of HMG CoA reductase in cholesterol metabolism and nonsterol isoprenoid biosynthesis, it is not surprising that it is one of the most highly regulated enzymes in nature. Inhibition of HMG CoA reductase by statins controls the production of cholesterol with benefits for lowering blood pressure, the treatment of

cardiovascular disease, and inflammatory processes. Lipitor, a statin, is the best-selling pharmaceutical in history with reported sales of more than $12.4 billion in 2008 *(32)*.

The DXP Pathway

For many years it was believed that the MVA pathway was responsible for the biosynthesis of isoprenoids in all organisms. However, several findings were inconsistent with this idea. For instance, labeling studies did not incorporate [^{13}C] acetate into ubiquinone in *E. coli (33)* or pentalenolactone in *S. chromofuschus (34,35*). Feeding experiments using [U-^{13}C] glucose and pentalenolactone-producing *S. exfoliatus* showed labeling patterns inconsistent with mevalonate-dependent synthesis *(34,35)*. Moreover, the specific HMG CoA reductase inhibitor mevinolin did not suppress *E. coli* growth *(33)*.

In the 1990s, Rohmer and Arigoni discovered the DXP pathway in *E. coli* (Figure 1.3) *(36)*, which begins with the thiamine-dependent condensation of pyruvate and glyceraldehyde-3-phosphate catalyzed by DXP synthase to produce deoxyxylulose phosphate (DXP) *(16, 37)*. DXP is then converted to methylerythritol-5-phosphate (MEP) in the presence of NADH by DXP reductoisomerase. This reaction is the first committed step in the DXP pathway. MEP is then linked to CTP by MEP cytidylyltransferase to form cytidylyldiphosphomethylerythritol (CDP-ME), which is then phosphorylated by CDP-MEP kinase to yield cytidylyldiphosphomethylerythritol phosphate (CDP-MEP). Cyclomethylerythritoldiphosphate (cMEPP) synthase then eliminates CMP from CDP-MEP to form the cyclic product cMEPP, which is then reduced by 4-hydroxydimethylallyl diphosphate (HDMAPP) synthase. The reductive dehydration of

HDMAPP is then catalyzed by HDMAPP reductase to produce an approximately 5:1 mixture of IPP and DMAPP *(38)*. In the DXP pathway, IPP isomerase is not required to produce DMAPP from IPP. However, organisms harboring the DXP pathway still utilize IPP isomerase to balance the pools of IPP and DMAPP depending on current metabolic needs *(39)*.

Possible evidence for feedback regulation of cMEPP synthase (also known as IspF) comes from crystallographic, NMR and mass spectrometric studies in which the enzyme is found to bind phosphate, IPP, DMAPP, geranyl diphosphate and farnesyl diphosphate, all of which are downstream products of the DXP pathway *(40,41)*. However, there is currently no evidence for the global transcription regulation of DXP pathway genes as they are dispersed throughout genomes, in contrast with MVA pathway genes in some microorganisms that are organized into operons that are likely transcription regulated. On the other hand, the genes for MEP cytidylyltransferase (IspD) and cMEPP synthase (IspF) are transcriptionally fused in some cases, coding for a bifunctional enzyme that can catalyze nonconsecutive steps in the pathway *(37)*. IspDF can then form complexes with IspE to create an assembly of catalytic centers that can efficiently catalyze the three consecutive steps from MEP to cMEPP at the heart of the DXP pathway. Nonetheless, no evidence exists for rate enhancement by complex formation or substrate channeling *(42,43)*.

DXP pathway enzymes are absent from mammals but present in important pathogens that cause malaria, tuberculosis and a variety of sexually transmitted diseases *(15)*. For example, DXP enzymes in *P. falciparum* translocate to the apicoplast, an organelle of cyanobacterial origin absent from mammals *(44,45)*. The substrates and

products of DXP enzymes are also unique and their activities cannot be compensated for by another enzyme *(46)*. These make them attractive targets for therapeutic intervention via specific drug inhibition. Fosmidomycin, for instance, is an inhibitor of DXP reductoisomerase and has been used effectively in combination with clindamycin for the treatment of malaria *(47-50)*. The presence of DXP enzymes in plant chloroplasts also suggest them as suitable targets for herbicide development *(37)*. Improved sources of secondary plant metabolites with medical importance such as artemisinin and taxol as well as other terpenoids used in essential oils and flavoring can also be developed by genetic manipulation of the DXP pathway *(51)*.

Archaeal Cell Membrane Lipids

The three-domain classification scheme for all living things first emerged in the 1970s with the discovery by Woese and coworkers of the third domain of life, called the Archaea *(52)*. The development of molecular classification techniques enabled comparisons of ubiquitous sequences such as ribosomal proteins and RNAs that strongly suggested the early divergence of the archaeal and bacterial domains from within the prokaryotic lineage *(53)*. Since then, a wealth of other findings obtained from comparisons of membrane lipid structures and genome sequences provided strong support to the existence of this third domain *(54-57)*. Moreover, contrary to the previous understanding that archaeal organisms only thrived in extreme environments (high salt, low pH, high temperature), it was found that they also inhabit milder terrestrial and aquatic habitats *(58,59)*.

Isoprenoid membrane lipids are the key features of Archaea *(60-65)*. Unlike the bacterial lipid bilayer composed of 16- to 18-carbon fatty acid side chains in ester linkages with glycerol, archaeal lipids contain ether-linked isoprenoid chains ranging in length from 20 to 40 carbon atoms. The most common archaeal membrane lipid is diphytanylglycerol diether (archaeol), which contains phytanyl chains consisting of 20 carbon atoms. Archaeal membrane lipids with acyl chains of different lengths (e.g. C25) have also been observed *(66,67)*. In another case, diphytanylglycerol tetraether (caldarchaeol) forms a lipid monolayer, where each lipid subunit consists of two hydrophilic groups connected by 40-carbon hydrophobic isoprene units *(68)*. Modifications of the isoprenoid chain include formation of macrocyclic diether lipids by joining terminal methyl groups, a bridge between the alkyl chains or a break in one of the alkyl chains in tetraether lipids, introduction of pentacyclic rings *(69)*, and addition of hydroxyl groups (Figure 1.4).

The composition of membrane lipids in various archaeal species can change in response to environmental factors. For example, in the hyperthermophilic archaea *Thermococcus kodakarensis* and *Archaeoglobus fulgidus*, increasing growth temperature leads to an increase in the caldarchaeaol to archaeol ratio *(70,71)*. In *Methanocaldococcus jannaschii*, archaeols decreased and caldarchaeols and macrocyclic archaeols increased with increasing growth temperature *(72)*. The number of cyclopentane rings also increases with increasing growth temperature in Thermoplasmatales and Sulfolobales *(73-75)*. Several other examples of membrane lipid adaptations in archaea are known, although the molecular basis of this thermosensing and the resulting changes in membrane lipids remain to be elucidated.

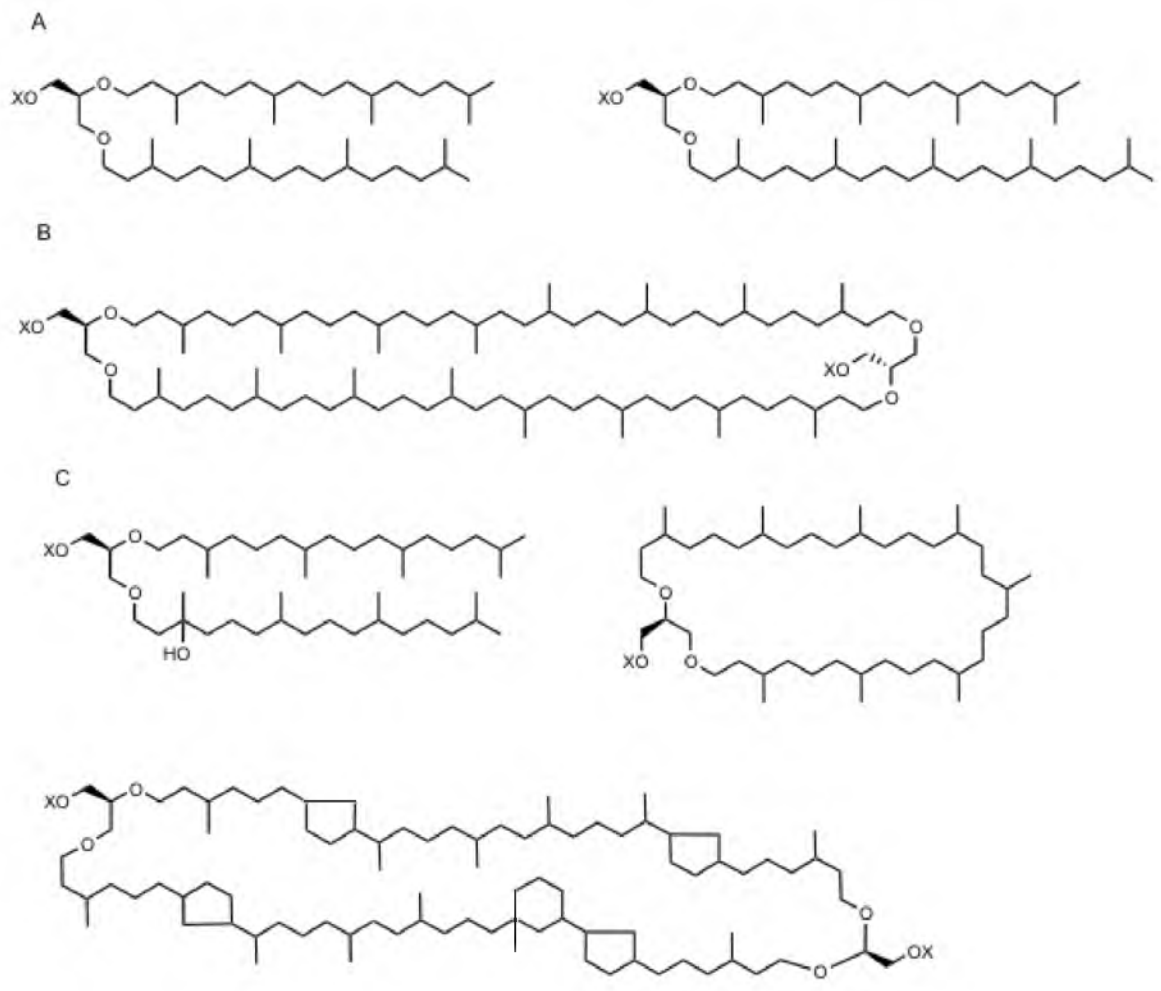


Figure 1.4. Examples of isoprenoid lipids found in Archaeal membranes. X represents sugars or other polar head groups. (a) *sn*-2,3-diacylglycerol-diether lipid (b) Glycerol-dialkyl-glycerol tetraether lipid (c) Common modifications of the isoprenoid side chain.

The differences between the archaeal and bacterial/eukaryotic membranes go beyond the structures of the lipid side chains. Whereas bacteria and eukaryotes utilize 1,2-*sn*-glycerol, archaea utilize the other stereoisomer, 2,3-*sn*-glycerol *(62, 65,76)*. This difference in stereochemistry is absolute with no known exception in any one of the three domains. As such, the ether-linked isoprenoid membrane lipids are the most clear-cut identifying features of Archaea *(76, 77)*, which are also generally accepted to confer strength, stability and limited permeability to their membranes. Ether-linked lipids can be considered more chemically stable at low pH and high temperatures than ester-linked lipids, and branched isoprenoid chains can form more densely packed supramolecular structures that are more stable and less permeable than bacterial/eukaryotic membranes.

Although there is ample experimental evidence for the idea that ether-linked isoprenoid membranes are well suited to support life in extreme environments *(78-80)*, there is no strict correlation between isoprenoid membranes and extremophilicity, as some archaea that live on nonextreme environments also harbor the same ether-linked isoprenoid lipids. Moreover, some thermophilic bacteria such as members of the hyperthermophilic Aquifex and Thermotoga genera are known to contain a majority of ester-linked fatty acid membrane lipids in addition to ether-linked ones *(81)*. In this case, the ether-linked lipids contain fatty acid side chains, not isoprenoids, with the same stereochemistry as bacterial/eukaryotic lipids.

The Archaeal Mevalonate Pathway

Of the seven enzymes required to synthesize IPP and DMAPP in eukaryotes, orthologs of only four have been identified in archaea, including acetoacetyl CoA

thiolase, HMG CoA synthase, HMG CoA reductase, and MK *(82)*. While orthologs of the eukaryotic type 1 IPP isomerase (IDI1) are generally absent from archaea except Halobacteriales, the nonorthologous enzyme that catalyzes the same reaction, type 2 IPP isomerase (IDI2), is found in all archaea including Halobacteriales *(82)*. Thus, two enzymes required for the conversion of phosphomevalonate to IPP – PMK and PPMD – are generally missing in archaea except in some restricted archaeal clades, perhaps as a result of horizontal gene transfer from eukaryotes or bacteria.

Within archaea, homologs of PMK and PPMD have only been identified in Sulfolobus *(82)*. These enzymes are likely of eukaryotic origin as shown by phylogenetic analyses. Thermoplasmatales and Halobacteriales also possess homologs of PPMD, although both cluster in phylogenetic trees with bacterial PPMDs and not with Sulfolobus PPMD *(82)*. Nonhomologous enzyme recruitment has been proposed by some authors to replace these missing enzymes in the archaeal MVA pathway, as is the case of the metazoan PMK similar to viral nucleoside monophosphate kinase instead of enzymes from the galactokinase, homoserine kinase, MK and PMK families *(83,84)*. Another notable example is that of the nonhomologous IDI2 mentioned earlier. First discovered in *Streptomyces sp.* by Kaneda *et al.*, this enzyme catalyzes the isomerization of IPP to DMAPP and is different from the mammalian IDI1 *(85)*. IDI2 was later found to be encoded in the genomes of many archaea and was characterized in several of them *(86,87)*. In addition, a search for enzymes in *Methanocaldococcus jannaschii* (MJ) able to phosphorylate phosphomevalonate led Grochowski *et al.* to identify a gene (MJ044) that codes for an enzyme with ATP-dependent kinase activity towards IP *(88)*. The enzyme was assigned as isopentenyl phosphate kinase (IPK). Its gene is found in all

archaea except Nanoarchaea and clusters with other genes in the MVA pathway. Moreover, the IPKs from MJ, THA, and MTH have high catalytic efficiencies that are consistent with those of established enzymes in the MVA pathway. These observations led to the proposition of an alternate route in the archaeal MVA pathway in which the two steps for the conversion of phosphomevalonate to IPP are the reverse of those in the classical MVA pathway. This alternate pathway requires that a phosphomevalonate decarboxylase (PMD) precedes the final phosphorylation step, and that the MJ043 gene, which also clusters with other MVA pathway genes in archaea, is proposed to code for this enzyme *(88)*. The decarboxylation reaction does not require a substrate with a diphosphate group so that the proposed alternate route can proceed from phosphomevalonate as the substrate of the putative PMD.

The Early Evolution of the Mevalonate Pathway in the Three Domains of Life

The dichotomy between the MVA pathway in eukaryotes and archaea and the DXP pathway in bacteria and plastid-bearing eukaryotes has been recognized for many years. However, the observations that archaea lacked the two enzymes used in the conversion of phosphomevalonate to IPP, as well as the presence of the MVA pathway in Firmicutes, raised some questions about the evolution of isoprenoid pathways in living things. Is the alternate MVA pathway conserved in archaea? Given the parsimonious nature of genomes and the identification of the MVA pathway in Firmicutes, is the MVA pathway rather than the DXP pathway ancestral to the bacterial domain? Phylogenetic studies performed by Lombard and Moreira showed that despite the complexity of the

distribution of isoprenoid biosynthetic pathways in the three domains of life, there exists a clear preference for the DXP pathway in bacteria, the classical MVA pathway in eukaryotes and the alternate MVA pathway in archaea *(89)*. Thus, instead of the two isoprenoid pathways distributed between the three domains and some cases of HGT among them, there is strong phylogenetic evidence for a tripartite distribution in which the alternate MVA pathway is considered a separate and conserved pathway in archaea. Despite these preferences, however, the same phylogenetic studies led Lombard and Moreira to hypothesize that the classical MVA pathway was likely present in the last universal common ancestor (LUCA) of all organisms and is therefore the ancestral isoprenoid pathway. A more detailed discussion of this analysis is presented below.

The Alternate MVA Pathway Is Conserved in Archaea

Phylogenetic analyses of MVA enzymes (except acetoacetyl CoA thiolase, which is not exclusive to this pathway) from a wide list of representative species showed that the alternate MVA pathway is characteristic of the whole archaeal domain *(89)*. The first three enzymes HMGS, HMGR and MK showed a general topology consistent with the accepted phylogeny of organisms. For example, HMGS yielded phylogenetic trees with a well-supported separation of the eukaryotic, bacterial and archaeal sequences down to the level of phyla or classes, with the exception of the bacterial phylum Chloroflexi and the bacteriodete *Flavobacterium johnsoniae* whose HMGS sequences branched among the archaea, indicating HGT from this domain *(89, Figure 2)*. The last three enzymes in the classical MVA pathway, PMK, PPMD and IDI1, did not exhibit this tridomain distribution since they are generally missing in archaea, with the exception of

Haloarchaea and Thermoplasmatales, which have PPMD, and Haloarchaea and Thermautoarchaeota, which have IDI1 (*89, Supplementary Figures 2-4*). These PPMD and IDI1 homologues branched among the bacteria, suggesting their HGT from the bacterial domain. Moreover, the only PMK and PPMD sequences found in archaea (Sulfolobales) did not branch with those of either eukayotes or bacteria, but were in an intermediate position that suggests their ancestral presence only in Sulfolobales. Orthologues of IPK and IDI2 were also found in archaea with the other MVA pathway enzymes, except *Sulfolobus tokodaii*, *Cenarchaeum symbiosum*, *Nitorospumilus maritimus* and some halophilic archaea, where IDI2 is replaced by IDI1. Phylogeny of IDI2 sequences showed a separation between bacterial and archaeal sequences, except for the Firmicutes, which branched among the archaea (*89, Supplementary Figure 5*). In addition, IPK homologues could not be detected in the genomes of most species in other domains of life, with the exception of 14 eukaryotes and 9 bacteria *(89)*. IPK sequences are less conserved across species compared to those of other MVA enzymes, leading to phylogenetic trees that were poorly supported and less congruent with the presumed phylogeny of organisms. However, removal of bacterial and eukaryotic IPK homologues led to better alignments and more supported trees with monophyletic archaeal classes. Thus, despite sporadic cases of HGT events, there is strong evidence of a conserved, modified MVA pathway for the biosynthesis of IPP and DMAPP in archaea *(89)*.

An Ancestral MVA Pathway in Bacteria

The current availability of a larger sampling of complete bacterial genome sequences has led to the new finding that the classical MVA pathway is ancestral to

bacteria despite the fact that the DXP pathway is now the dominant route for isoprenoid biosynthesis in this domain. MVA pathway enzymes have been identified not only in Firmicutes, but also in 93 bacterial species belonging to 43 different genera of 6 phyla *(89)*. In particular, Firmicutes, Actinobacteria, Proteobacteria and the Spirochete genus Borrelia encode most of the enzymes in the MVA pathway. Bacteriodetes possess only some of the enzymes, while Chloroflexi contains IPK instead of PMK. IDI1 and IDI2, compulsory in the MVA pathway of eukaryotes and archaea, have been identified in bacteria, despite the presence of HDMAPP reductase, which is able to synthesize both IPP and DMAPP simultaneously. While it has been shown that in certain bacteria the IDIs complement defects in the synthesis of IPP in Fe-S protein biogenesis *(90)*, their major role in Firmicutes is in isoprenoid biosynthesis. Bacteriodetes, some Actinobacteria, and some Proteobacteria employ IDI1, whereas Firmicutes, Cyanobacteria, Chlorobi, Chloroflexi, Deinococcus-thermus, Spirochetes, Deltaproteobacteria, and some proteobacterial and archaeobacterial species utilize IDI2. In addition, some bacteria contain either one of two classes of HMGR. Class II HMGR is more widespread and characteristic of many bacterial phyla and is considered ancestral in this domain. The Class I homologue is found in several Proteobacteria and Actinobacteria.

In phylogenetic analyses, the bacterial sequences of MVA pathway enzymes emerged as a monophyletic group distinct from the archaeal and eukaryotic clades. In fact, for each enzyme, the vast majority of bacterial sequences form an independent monophyletic group, as in the phylogenies of HMGS *(89, Figure 2)* and MK *(89, Figure 4)*. Biochemical characterizations of these enzymes also showed that they have distinct

sequence signatures and biochemical properties. These indicate that contrary to the commonly accepted idea that bacteria obtained MVA pathway enzymes by HGT from archaeal or eukaryotic donors, the MVA pathway is ancestral to bacteria. Only in the cases of HMGS from Chloroflexi *(89, Figure 2)* and *Flavobacterium johnsoniae*, and IDI2 in Firmicutes that branch with the archaeal clade *(89, Supplementary Figure 5)* was the expected pattern of HGT observed, indicating that these bacterial organisms obtained some of their MVA enzymes from archaeal donors. Conversely, the Class II HMGR of euryarchaeota and thaumarchaeota clustered within the bacterial clade, signifying HGT from bacteria to archaea *(89, Figure 3)*.

The apparent separation of bacterial MVA proteins from those of archaea and eukaryotes can be explained in three scenarios *(89)*. First, an ancient bacterial lineage may have obtained the MVA pathway enzymes from an archaeal or bacterial donor, followed by a period of fast evolution that led to the high sequence divergence of the bacterial MVA enzymes from the original ones. This was followed by new HGT events that led to the presence of MVA pathway enzymes in today's bacterial classes (Figure 1.5a). The second scenario involves the inheritance of the entire MVA pathway from archaea or eukaryotes prior to the speciation of bacteria. In this case, the MVA pathway becomes ancestral to bacteria (Figure 1.5b). The third scenario assumes the presence of the MVA pathway from the cenancestor, so that vertical inheritance led eukaryotes, archaea and bacteria to possess an ancestral MVA pathway prior to the divergence of their respective sequences (Figure 1.5c). The last two scenarios require multifarious losses of MVA pathway genes from different bacterial lineages made possible by the

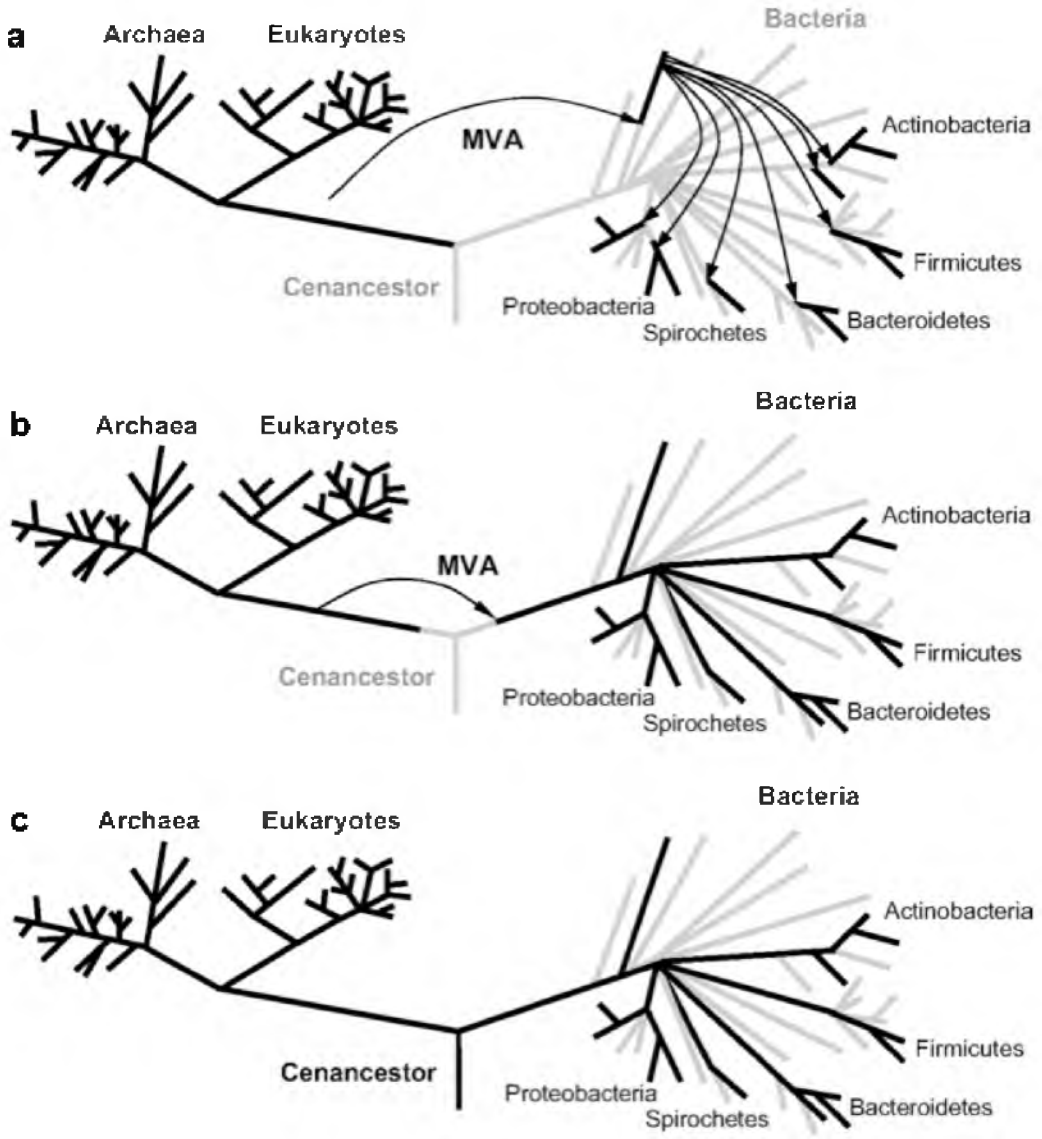


Figure 1.5. Three hypotheses for the origin of MVA pathway enzymes in bacteria. (a) Horizontal transfer of the complete MVA pathway from eukaryotic or archaeal donors to a bacterial lineage and subsequent transfers from this lineage to bacterial groups currently utilizing the MVA enzymes (black branches). (b) Ancient horizontal transfer of the entire MVA pathway to a bacterial ancestor prior to the divergence of the contemporary bacterial phyla. (c) The cenancestor as the source of the MVA pathway present in some bacterial species. Scenarios (b) and (c) implies multiple losses of the MVA pathway in some bacterial phyla (grey branches). The classical rooting of the tree of life is used for convenience.

functional redundancy of the MVA pathway with the DXP pathway present in most bacteria.

An Ancestral MVA Pathway in the Cenancestor

While there is a clear preference for the MEP pathway in bacteria, the classical MVA pathway in eukaryotes, and the alternative MVA pathway in archaea, phylogenetic analyses of MVA pathway enzymes by Lombard and Moreira support the possible presence of the classical MVA pathway in the cenancestor, subsequently inherited by each of the three domains of life *(89)*. This hypothesis entails the massive loss of the classical MVA pathway enzymes in bacteria and its replacement by the MEP pathway, and in archaea, the early presence of the missing PMK and PPMD enzymes (Figure 1.6a). Thus, the current repertoire of enzymes in both bacterial and archaeal domains is the result of both small-scale and large-scale nonhomologous replacement of enzymes in the classical MVA pathway. Still, two alternative hypotheses can explain the evolution of isoprenoid biosynthesis pathways. On the one hand, it is possible that the cenancestor had ancestral forms of each of the three different isoprenoid biosynthetic pathways, followed by selective losses in each domain of life to account for their unique differences (Figure 1.6b). Worthy of note is the absence of MEP pathway enzymes in archaea and in nonplastid bearing eukaryotes, signifying that the common ancestor of these two domains did not possess the MEP pathway. On the other hand, the cenancestor could also not have possessed any of the isoprenoid pathway enzymes and was therefore acellular *(91,92)* or devoid of a lipid membrane *(93)* (Figure 1.6c). This would entail that each isoprenoid pathway arose independently in each domain, with complex exchanges of enzymes

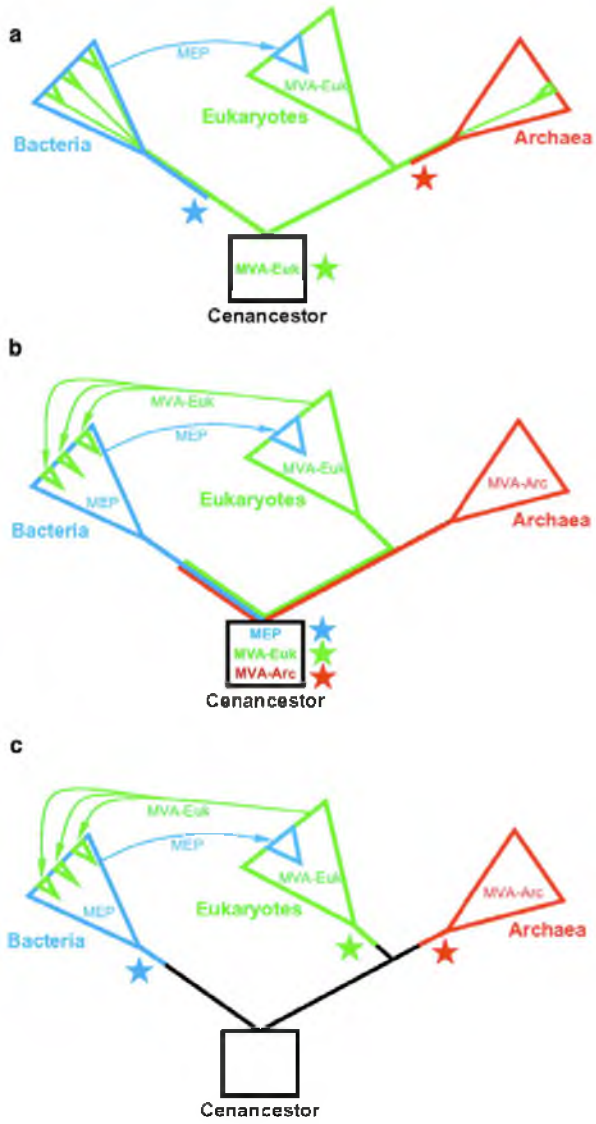


Figure 1.6. Scenarios that explain the evolution of the different isoprenoid biosynthesis pathways on a schematic three domains of life tree. The eukaryotic and archaeal MVA pathways (MVA-Euk and MVA-Arc) are indicated in green and red, respectively, while the bacterial DXP pathway (here indicated as MEP) is indicated in blue. Stars indicate the origin of the respective pathways. (a) The classical MVA pathway would have been present in the cenancestor and supplanted in most bacteria and archaea by the MEP and the MVA-Arc pathways, respectively. (b) The three pathways would have been possessed by the cenancestor, but only one pathway would have been conserved in each domain of life. HGT events would explain the occurrence of supplementary pathways. (c) The cenancestor did not possess any of the three isoprenoid pathways, and each would have appeared in the ancestors of each domain life. Subsequent HGT events would explain the presence of MVA-Euk and MEP pathways in several bacteria and eukaryotes, respectively.

between the bacterial and eukaryotic domains. This hypothesis has been advanced by some authors, although it is remote because of the presence of fatty acid biosynthesis and degradation enzymes in archaea, as well as of dehydrogenases involved in glycerol-1-phosphate (G1P) and glycerol-3-phosphate (G3P) syntheses in archaea and bacteria, which argue that all the components required to create a lipid membrane (phospholipids composed of glycerol phosphate bound to lateral chains of isoprenoids or fatty acids) existed in the cenancestor *(94)*. In other words, the differences in cellular membranes observed today are likely secondary and related to the divergence of the three domains of life.

Isopentenyl Phosphate Kinase

Isopentenyl phosphate kinase (IPK) is one of the missing enzymes in the alternate MVA pathway first discovered in MJ in 2006 by Grochowski *et al. (88)*. Since then its homologues in THA and MTH have been kinetically characterized, including its substrate specificity, product formation, forward and reverse reactions, and pH- and temperature-activity profiles *(95)*. IPK is the first enzyme found to catalyze the ATP-dependent phosphorylation of IP to form IPP *in vitro (88)*. Although other enzymes in MJ display preference for GTP, MJ IPK strictly uses ATP. MJ IPK has an apparent K_M^{IP} of 256 μM, with a k_{cat} of 3.3 x 10^2 s^{-1}. This K_M^{IP} is in line with the values measured for MK from rats, humans and MJ, which range from a $K_M^{R,S\text{-}mevalonate}$ of 68 to 288 μM. For MTH and THA, K_M^{IP} are 12.7 μM and 4.4 μM, respectively. All three enzymes have catalytic efficiencies comparable to other enzymes in the MVA pathway, suggesting their true roles as IP kinases (Table 1.1). Moreover, the phosphoryl transfer reaction has an equilibrium

Table 1.1. Published kinetic constants of IPKs from MJ, MTH and THA.

Constant	MJ IPK	MTH IPK	THA IPK
k_{cat} (s^{-1})	3.3×10^2	27.5 (± 0.3)	8.0 (± 0.2)
K_M^{IP} (μM)	256	12.7 (± 0.6)	4.4 (± 0.5)
K_M^{ATP} (μM)	*not determined*	13.4 (± 0.8)	6.0 (± 0.5)
K_d^{IP} (μM)	*not determined*	15.3 (± 2.6)	4.6 (± 1.5)
K_d^{ATP} (μM)	*not determined*	16.2 (± 2.5)	6.3 (± 2.2)
k_{cat}/K_M^{IP} (M^{-1} s^{-1})	1.3×10^6	2.2×10^6	1.3×10^6

constant K_{eq} = 6.3, indicating that *in vitro* the forward reaction is only slightly preferred than the reverse reaction *(95)*.

Published studies on IPKs confirmed their expected thermostability, with MJ IPK retaining 38% of its activity after incubation at 100 °C for 10 min, and MTH and THA IPKs remaining stable up to 50 °C and 70 °C, respectively. MJ IPK is active over a broad pH range between 7 and 9. MTH IPK prefers alkaline conditions while THA IPK prefers acidic conditions. The k_{cat}^{IP} values for MTH IPK are independent of pH from 7 to 9, which suggests a single rate-limiting step with the same ionization state throughout the tested pH range. In addition, the k_{cat}^{IP}/K_M^{IP}-pH profile for MTH IPK is bell-shaped, with two inflection points at pK_1 = 6.7 for the acidic limb and pK_2 = 9.9 for the basic limb. The pK_1 likely corresponds to the ionization of the second proton of IP or of a conserved histidine (His52). The pK_2 of 9.9 may correspond to the ionization of any one of the conserved lysine residues in the active site (Lys5, Lys14 and Lys223, pK_2 = 8.9 and pK_3 = 10.20) whose pK_a value was perturbed by neighboring interactions in the active site. Likewise, the k_{cat}^{IP}/K_M^{IP}-pH activity profile for THA IPK is bell-shaped, with a pK_1 for the acidic limb of 6.5 similar to that of MTH IPK, and pK_2 for the basic limb of 8.5 that may also reflect the ionization of a conserved active site lysine residue (Lys5, Lys14 and Lys205). The k_{cat}^{IP}-pH profile for THA IPK is also bell-shaped, suggesting multiple ionization states for the enzyme-substrate complex within the experimental pH range of 5.5 to 9.0. In addition, kinetic analyses by Chen and Poulter showed that the binding of the substrates IP and ATP is sequential, although further experiments need to be performed in order to determine their binding order *(95)*.

A blastp search against the NCBI nonredundant protein database using the MJ IPK sequence as a query retrieved 107 enzymes belonging to the amino acid kinase (AAK) family, of which 78 are from *Archaea* (representing 21 different families within the domain), 21 from *Eukarya* and 8 from *Bacteria*. Among these, 5 are annotated as IP kinases and 2 as fosfomycin kinases *(95)*. Enzymes in the AAK family catalyze the ATP-dependent phosphorylation of small molecules that contain phosphate, phosphonate, carboxylate and carbamate moieties that serve as the attacking nucleophile in the phosphoryl transfer reaction. MTH IPK has substantial sequence similarity with MJ IPK (*e*-value of $5e^{-29}$), while THA IPK was the most distant in sequence (*e*-value of $7e^{-13}$) among the retrieved homologues. Among the sequences are two fosfomycin kinases (FomA) from *Streptomyces wedmorensis* and *Streptomyces fradiae*, which are bacterial resistance enzymes capable of inactivating fosfomycin by the covalent modification of its phosphonate group. The crystal structure of *S. wedmorensis* FomA was recently determined and confirmed the putative binding and catalytic roles of several residues conserved in the alignment *(96,97)*. Like FomA, IPK from either THA or MTH is dimeric and this oligomerization is presumed to be important for the stabilization of loops within the vicinity of the IP binding site. Archaeal IPK is not known to interact with another enzyme and its product IPP is relatively stable, precluding the use of dimerization for substrate channeling.

Alignment of the retrieved sequences showed that most of the residues known to perform binding and catalytic roles in well-studied AAK enzymes (*e.g.*, N-acetylglutamate kinase or NAGK, *98*) are partially or fully conserved in putative IPKs (*95*, *Table S2*). Worthy of note are two conserved residues, Lys14 and His50 in THA

IPK, that are absolutely conserved in putative archaeal IPKs and in FomA but not in other AAK enzymes (Figure 1.7). The corresponding histidine residue in *S. wedmorensis* FomA (His58) is seen in the crystal structure to interact with a phosphonate oxygen atom of fosfomycin through water-mediated hydrogen bonding, suggesting that it may have a role in the orientation of the substrate for nucleophilic attack on the donor nucleotide. In addition, His58 is located in the αB helix, a secondary structure seen only in FomA and UMPK (which does not have a corresponding histidine residue) *(99)* but not in other enzymes belonging to the AAK family. This residue may therefore be characteristic of IPKs and FomAs and may serve as a fingerprint for identifying homologs. This commonality between archaeal IPKs and bacterial FomAs may also suggest an early evolutionary relationship between the two enzymes. Finally, a lysine residue (Lys14 in THA) is conserved in putative IPK and FomA homologs *(99)*. Recent structures of FomA as well as mutagenesis experiments showed that this conserved lysine residue is important for transition state stabilization of the negatively charged transition state *(100)*.

MTH and THA IPKs have also been shown to efficiently phosphorylate a variety of C4 and C5 organic phosphates with k_{cat} values comparable to the native substrate IP and slightly higher K_M values in the μM range. For the C10 substrate geranyl phosphate, the k_{cat} values are significantly lower (200-fold and 2000-fold lower for MTH and THA IPK, respectively) compared to native substrate, while the K_M values are in the high μM to low mM range. The catalytic efficiencies of MTH and THA IPK for the phosphorylation of GP are, however, both similarly low *(95)*.

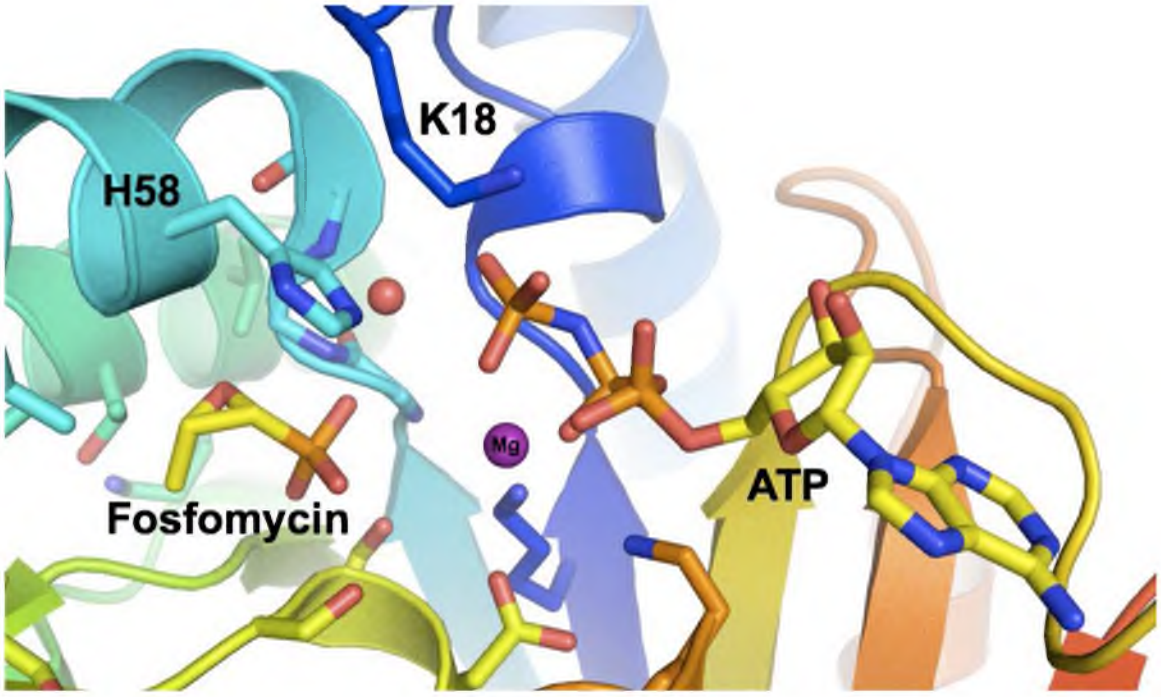


Figure 1.7. Active site of *Streptomyces wedmorensis* FomA, showing His58 and Lys14 that are conserved only in FomA and putative IPKs.

Thesis Overview

So far, only a few profiling studies on IPK have been documented, including the kinetic characterizations briefly discussed above. While more kinetic experiments need to be performed in order to fully describe the catalytic mechanism of IPK (*e.g.*, substrate binding order, substrate and product inhibition, cooperative binding kinetics observed in other AAK enzymes such as aspartokinase, mutagenesis studies to elucidate the roles of active site residues *etc.*), of utmost importance are three-dimensional structures of this enzyme that can reveal the residues important for substrate binding and catalysis. These structures will also locate residues found in sequence alignments to be partially or fully conserved, thus revealing the similarities and more importantly, the differences of IPK structures from those of other AAK enzymes whose three-dimensional structures have been determined. Finally, IPK structures can guide the structure-based engineering of the active site to create "new enzymes" or mutants capable of phosphorylating a variety of longer chain aliphatic or isoprenoid phosphates, such as geranyl phosphate or farnesyl phosphate and their analogues. These mutants can then be used as a tool for the radioactive labeling of the said longer-chain monophosphates that have potential applications in the studies of enzymes that use them as substrates, such as prenyltransferases and isoprenoid synthases.

The work described in this dissertation focuses on the crystal structure, substrate promiscuity and structure-based engineering of IPK. Chapter 2 describes the efforts in crystal structure determination of both MTH and THA IPK, in complex with their substrates and products. Chapter 3 then describes the studies done on the promiscuity of THA IPK for the antibiotic fosfomycin, where the results point toward a possible

evolutionary relationship between IPK and FomA initiated by an ancient HGT event, a hypothesis first proposed by Chen and Poulter *(95)*. Finally, Chapter 4 describes several THA IPK variants with apparent kinase activities toward the longer chain substrates, GP and FP.

References

1. Reiling, K. K., Yoshikuni, Y., Martin, V. J. J., Newman, J., Bohlmann, J., and Keasling, J. D. (2004) Mono and diterpene production in *Escherichia coli*, *Biotechnol. Bioeng. 87*, 200-212.

2. Sobotnik, J., Hanus, R., Kalinova, B., Piskorski, R., Cvacka, J., Bourguignon, T., and Roisin, Y. (2008) *(E,E)*-α-Farnesene, an alarm pheromone of the termite *Prorhinotermes canalifrons, Journal of Chemical Ecology 34 (4)*, 476-486.

3. Hern, A., and Dorn, S. (1999) Sexual dimorphism in the olfactory orientation of adult *Cydia pomonella* in response to α-farnesene, *Entomologia Experimentalis et Applicata 92 (1)*, 63-72.

4. Gibson, R. W., and Pickett, J. A. (1983) Wild potato repels aphids by release of aphid alarm pheromone, *Nature 302*, 608-609.

5. Wang, S., Tseng, T., Huanga, C., and Tsaic, T. (2004) Gardenia herbal constituents: applicable separation procedures, *Journal of Chromatography B 812*, 193-202.

6. White, N. J. (1997) Assessment of the pharmacodynamic properties of antimalarial drugs in vivo, *Antimicrob. Agents Chemother. 41*, 1413-1422.

7. Douglas, N. M., Anstey, N. M., Angus, B. J., Nosten, F., and Price, R. N. (2010) Artemisinin combination therapy for vivax malaria, *Lancet Infect Dis 10*, 405-416.

8. Fuchs, D. A. and Johnson, R. K. (1978) Cytologic evidence that taxol, an antineoplastic agent from *Taxus brevifolia*, acts as a mitotic spindle poison, *Cancer Treatment Reports 62*, 1219-1222.

9. Saville, M., Lietzau, J., Pluda, J., Feuerstein, I., Odom, J., Wilson, W., Humphrey, R., Feigal, E., Steinberg, S., and Broder, S. (1995) Treatment of HIV-associated Kaposi's sarcoma with paclitaxel, *Lancet 346*, 26-28.

10. Jahn, U., Simmons, R., Sturt, H., Grosjean, H., and Huber, H. (2004) Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp strain KIN4/I, *Arch. Microbiol 182*, 404-413.

11. Goldstein, J. L., and Brown, M. S. (1990) Regulation of the mevalonate pathway, *Nature 343*, 425-430.

12. Poulter, C. D. (2006) Farnesyl diphosphate synthase. A paradigm for understanding structure and function relationships in *E*-polyprenyl diphosphate synthases, *Phytochem. Rev. 5*, 17-26.

13. Davis, E. M. and Croteau, R. (2000) Cyclization enzyme in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes, *Top. Curr. Chem. 209*, 53-95.

14. Song, L., and Poulter, C. D. (1994) Yeast farnesyl diphosphate synthase: Site-directed mutagenesis of residues in highly conserved prenyltransferase domains I and II, *Proc. Nat. Acad. Sci. USA 91*, 3044-3048.

15. Rohdich, F., Bacher, A., and Eisenreich, W. (2004) Perspectives in anti-infective drug design. The later steps in the biosynthesis of the universal terpenoid precursors, isopentenyl diphosphate and dimethylallyll diphosphate, *Bioorg. Chem 32*, 292-308.

16. Kuzuyama, T. (2002) Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units, *Biosci. Biotechnol. Biochem. 66*, 1619-1627.

17. Brown, M. S. and Goldstein. J. L. (1980) Multivalent feedback regulation of HMG CoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth, *J. Lipid. Res. 21*, 505-517.

18. Goldstein, J. L. and Brown, M. S. (1984) Progress in understanding the LDL receptor and HMG CoA reductase, two membrane proteins that regulate the plasma cholesterol, *J. Lipid. Res. 25*, 1450-1461.

19. Osborne, T. F., Goldstein, J. L, and Brown, M. S. (1985) 5'-end of HMG CoA reductase gene contains sequences responsible for cholesterol-mediated inhibition of transcription, *Cell 42*, 203-212.

20. Sudhof, T. C., Russell, D. W., Brown, M. S. and Goldstein, J. L. (1987) 42 bp element from LDL receptor gene confers end-product repression by sterols when inserted into viral TK promoter, *Cell 48*, 1061-1069.

21. Sudhof, T. C., can der Westhuyzen, D. R., Goldstein, J. L., Brown, M. S., and Russell, D. W. (1987) Three direct repeats and a TATA-like sequence are required for regulated expression of the human low density lipoprotein receptor

gene, *J. Biol. Chem. 262*, 10773-10779.

22. Dawson, P. A., Hoffman, S. L., van der Westhuyzen, D. R., Sudhof, T. C., Brown, M. S., and Goldstein, J. L (1988) Sterol-dependent repression of low density lipoprotein receptor promoter mediated by 16-base pair sequence adjacent to binding site for transcription factor Sp1, *J. Biol. Chem. 263*, 3372-3379.

23. Osborne, T. F., Gil, G., Goldstein, J. L., and Brown, M. S. (1988) Operator constitutive mutation of 3-hydroxy-3-methylglutaryl coenzyme A reductase promoter abolishes protein binding to sterol regulatory element, *J. Biol. Chem. 263*, 3380-3387.

24. Smith, J. R., Osborne, T. F., Brown, M. S., Brown, M. S., Goldstein, J. L., and Gil, G. (1988) Multiple sterol regulatory elements in promotor for hamster 3-hydroxy-3-methylglutaryl-coenzyme A synthase, *J Biol. Chem. 263*, 18480-18487.

25. Nakanishi, M., Goldstein, J. L., and Brown, M. S. (1988) Multivalent control of 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Mevalonate-derived product inhibits translation of mRNA and accelerates degradation of enzyme, *J. Biol. Chem. 263*, 8929-8937.

26. Panini, S. R., Schnitzer-Polokoff, R., Spencer, T. A., and Sinensky, M. (1989) Sterol-independent regulation of 3-hydroxy-3-methylglutaryl-CoA reductase by mevalonate in Chinese hamster ovary cells. Magnitude and specificity, *J. Biol. Chem. 264*, 11044-11052.

27. Chin, D. J., Luskey, K. L., Anderson, R. G., Faust, J. R., Goldstein, J. L., and Brown, M. S. (1982) Appearance of crystalloid endoplasmic reticulum in compactin-resistant Chinese hamster cells with a 500-fold increase in 3-hydroxy-3-methylglutaryl-coenzyme A reductase, *Proc. Nat. Acad. Sci. USA 79*, 1185-1189.

28. Luskey, K. L., Faust, J. R., Chin, D. J., Brown, M. S. and Goldstein, J. L. (1983) Amplification of the gene for 3-hydroxy-3-methylglutaryl coenzyme A reductase, but not for the 53-kDa protein, in UT-1 cells, *J. Biol. Chem. 258*, 8462-8469.

29. Gibson, D. M. in Regulation of HMG CoA Reductase. (ed. Priess, B.) 80-132 (Academic, New York, 1985).

30. Brown, M. S., Brunschede, G. Y., and Goldstein, J. L. (1975) Inactivation of 3-hydroxy-3-methylglutaryl coenzyme A reductase in vitro, An adenine nucleotide-dependent reaction catalyzed by a factor in human fibroblasts, *J. Biol. Chem. 250*, 2502-2509.

31. Hardie, D. G., Carling, D., and Sim, A. T. R. (1989) The AMP-activated protein

kinase: a multisubstrate regulator of lipid metabolism, *Trends Biochem. Sci. 14*, 20-23.

32. Pfizer 2008 Annual Report. Pfizer. 23 April 2009.

33. Zhou, D., and White, R. H. (1991) Early steps of isoprenoid biosynthesis in *Escherichia coli, Biochem. J. 273*, 627-634.

34. Cane, D. E., Rossi, T., and Pachlatko, J. P. (1979) The biosynthesis of pentalenolactone, *Tetrahedron Lett. 20*, 3639-3642.

35. Cane, D. E., Rossi, T., Tillman, A. M., and Pachlatko, J. P. (1981) Stereochemical studies of isoprenoid biosynthesis: Biosynthesis of pentalenolactone from [U-$^{13}C_6$]glucose and [6-$^{2}H_2$]]glucose, *J. Am. Chem. Soc. 103*, 1838-1843.

36. Eisenreich, W., Bacher, A., Arigoni, D., and Rohdich, F. (2004) Biosynthesis of isoprenoids via the non-mevalonate pathway, *Cell. Mol. Life Sci. 61*, 1401-1426.

37. Hunter, W. N. (2007) The non-mevalonbate pathway of isoprenoid precursor biosynthesis, *J. Biol. Chem. 282*, 21573-21577.

38. Adam, P., Hecht, S., Eisenreich, W., Kaiser, J., Grawert, T., Arigoni, D., Bacher, A., and Rohdich, F. (2002) Biosynthesis of terpenes: studies of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate reductase, *Proc. Natl. Acad. Sci. USA 99*, 12108-12113.

39. Rothman, S. C., Helm, T. R., and Poulter, C. D. (2007) Kinetic and spectroscopic characterization of type II isopentenyl diphosphate isomerase from *Thermus thermophilus*: Evidence for formation of substrate-induced flavin species, *Biochemistry 46*, 5437-5445.

40. Ni., S., Robinson, H., Marsing, G. C., Bussiere, D. E., and Kennedy, M. A. (2004) Structure of 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase from *Shewanella oneidensis* at 1.6 Å: identification of a farnesyl pyrophosphate trapped in a hydrophobic cavity, *Acta Crystallogr. Sect. D Biol. Crystallogr. 60*, 1949-1957.

41. Kemp, L. E., Alphey, M. S., Bond, C. S. Ferguson, M. A., Hecht, S. A., Bacher, A., Eisenreich, W., Rohdich, F., and Hunter, W. N. (2005) The identification of isoprenoids that bind in the intersubunit cavity of *Escherichia coli* 2C-methyl-D-eryhthritol-2,4-cyclodiphosphate synthase by complementary biophysical methods, *Acta Crystallogr. Sect D. Biol. Crystallogr. 61*, 45-52.

42. Gabrielsen, M., Rohdich, F., Eisenreich, W., Grawert, T., Hecht, S., Bacher, A., and Hunter, W. N. (2004) Biosynthesis of isoprenoids: a bifunctional IspDF enzyme from *Campylobacter jejuni*, *Eur. J. Biochem. 271*, 3028-3035.

43. Lherbet, C., Pojer, F., Richard, S. B., Noel, J. P., and Poulter, C. D. (2006) Absence of substrate channeling between active sites in the *Agrobacterium tumefaciens* ISpDF and IspE enzymes of the methyl erythritol phosphate pathway, *Biochemistry 45*, 3548-3553.

44. Cassera, M. B., Gozzo, F. C., D'Alexandri, F. L., Merino, E. F., del Portillo, H. A., Peres, V. J., Almeida, I. C., Eberlin, M. N., Wunderlich, G., Wiesner, J., Jomaa, H., Kimura, E. A. and Katzin, A. M. (2004) The methylerythritol phosphate pathway is functionally active in all intraerythrocytic stages of *Plasmodium falciparum*, *J. Biol. Chem. 279*, 51749-51759.

45. Ralph, S. A., van Dooren, G. G.,Waller, R. F. Crawford, M. J., Fraunholz, M. J., Foth, B. J., Tonkin, C. J., Roos, D. S., and McFadden, G. I. (2004) Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast, *Nat. Rev. Microbiol. 2*, 203-216.

46. Hasan, S., Daugelat, S., Rao, P. S. S., and Schreiber, M. (2006) Prioritizing genomic drug targets in pathogens: application to Mycobacterium tuberculosis, *PLoS Comput. Biol. 6*, 1-12.

47. Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., Soldati, D., and Beck, E. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs, *Science 285*, 1573-1576.

48. Borrmann, S., Adegnika, A. A., Matsiegui, P. B., Issifou, S., Schindler, A., Mawili-Mboumba, D. P., Baranek, T., Wiesner, J., Jomaa, H., and Kremsner, P. G. (2006) Fosmidomycin-clindamycin for Plasmodium falciparum infections in African children, *J. Infect. Dis.189*, 901-908.

49. Bormann, S., Lundgren, I., Oyakhirome, S., Impouma, B., Matsiegui, P. B., Adegnika, A. A., Issifou, S., Kun, J. F., Hutchinson, D., Wiesner, J., Jomaa, H., and Kremsner, P. G.(2006) Fosmidomycin plus clindamycin for treatment of pediatric patients aged 1 to 14 years with Plasmodium falciparum malaria, *Antimicrob Agents Chemother. 50 (8)*, 2713-2718.

50. Ruangweerayut, R., Looareesuwan, S., Hutchinson, D., Chauemung, A., Banmairuroi, V., and Na-Bangchang, K. (2008) Assessment of the pharmacokinetics and dynamics of two combination regimens of fosmidomycin-clindamycin in patients with acute uncomplicated falciparum malaria, *Malaria J. 7(1)*, 225.

51. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids, *Nature Biotechnol. 21*, 796-802.

52. Forterre, P., Gribaldo, S., and Brochier -Armanet, C. in Natural History of the Archaeal Domain in Archaea: Evolution, Physiology and Molecular Biology. (eds. Garrett, R. A., and Klenk, H-P.) 39-50 (Blackwell Publishing, 2007).

53. Forterre, P., Brochier, C., and Philippe, H. (2002) Evolution of the *Archaea*. *Theor. Popul. Biol. 61*, 409-422.

54. Koga, Y., and Morii, H. (2005) Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects, *Biosci. Biotechnol. Biochem. 69*, 2019-2034.

55. Barry, E. R., and Bell, S. D. (2006) DNA replication in the archaea, *Microbiol. Mol. Biol. Rev. 70*, 876-887.

56. Werner, F. (2007) Structure and function of archaeal RNA polymerases, *Mol. Microbiol. 65*, 1395-1404.

57. Snel, B., Bork, P., and Huynen, M. A. (1999) Genome phylogeny based on gene content, *Nature Genet. 21*, 108-110.

58. Pace, N. R. (1997) A molecular view of microbial diversity and the biosphere, *Science 276*, 734-740.

59. Schleper, C., Jurgens, G., and Jonuscheit, M. (2005) Genomic studies of uncultivated archaea, *Nature Rev. Microbiol. 3*, 479-488.

60. Kates, M., Sastry, P. S., and Yengoyan, L. S. (1963) Isolation and characterization of a diether analog of phosphatidyl glycerophosphate from *Halobacterium cutirubrum*, *Biochim. Biophys. Acta. 70*, 705-707.

61. Kates, M., Palameta, B., Joo, C. N., Kushner, D. J., and Gibbons, N. E. (1966) Aliphatic diether analogs of glycerol-derived lipids. IV. The occurence of di-*O*-dihydrophytylglycerol ether containing lipids in extremely halophilic bacteria, *Biochemistry 5*, 4092-4099.

62. Kates, M. (1993) Membrane lipids of extreme halophiles: biosynthesis, function and evolutionary significance, *Experientia 49*, 1027-1036.

63. Sehgal, S. N., Kates, M., and Gibbons, N. E. (1962) Lipids of *Halobacterium cutirubrum*, *Can. J. Biochem. Physiol. 40*, 69-81.

64. De Rosa, M., Gambacorta, A., Nicolaus, B., Ross, H. N. M., Grant, W. D., and Bu'Lock, J. D. (1982) An asymmetric archaebacterial diether lipid from alkaliphilic halophiles, *J. Gen. Microbiol. 128*, 343-348.

65. Koga, Y., Nishihara, M., Morii, H., and Akagawa-Matsushita, M. (1993) Ether

polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses, *Microbiol. Rev. 57*, 164-182.

66. De Rosa, M., Gambacorta, A., and Gliozzi, A. (1986) Structure, biosynthesis, and physicochemical properties of archaebacterial lipids, *Microbiol. Rev. 50*, 70-80.

67. Morii, H., Yagi, H., Akutsu, H., Nomura, N., Sako, Y., and Koga, Y. (1999) A novel phosphoglycolipid archaetidyl(glucosyl)inositol with two sesterterpanyl chains from the aerobic hyperthermophilic archaeon *Aeropyrum pernix* K1, *Biochim. Biophys. Acta 1436*, 426-436.

68. Langworthy, T. A. (1977) Long-chain diglycerol tetraethers from *Thermoplasma acidophilum. Biochim. Biophys. Acta 487*, 37-50.

69. Gulik, A., Luzzati, V., De Rosa, M., and Gambacorta, A. (1985) Structure and polymorphism of bipolar isopranyl ether lipids from archaebacteria, *J. Mol. Biol. 182*, 131-149.

70. Lai, D., Springstead, J. R., and Monbouquette, H. G. (2008) Effect of growth temperature on ether lipid biochemistry in *Archaeoglobus fulgidus*, *Extremophiles 12*, 271-278.

71. Matsuno, Y., Sugai, A., Higashibata, H., Fukuda, W., Ueda, K., Uda, I., Sato, I., Itoh, T., Imanaka, T., and Fujowara, S. (2009) Effect of growth temperature and growth phase on the lipid composition of the archaeal membrane from *Thermococcus kodakarensis*, *Biosci. Biotechnol. Biochem. 73*, 104-108.

72. Sprott, G. D., Meloche, M., and Richards, J. C. (1991) Proportions of diether, macrocyclic diether, and tetraether lipids in *Methanococcus jannaschii* grown at different temperatures, *J. Bacteriol. 173*, 3907-3910.

73. De Rosa, M., Esposito, E., Gambacorta, A., Nicolaus, B., Bu'Lock, J. D. (1980) Effects of temperature on ether lipid composition of *Caldariella acidophila, Phytochemistry 19*, 827-831.

74. Shimada, H., Nemoto, N., Shida, Y., Oshima, T., and Yamagashi, A. (2008) Effects of pH and temperature on the composition of polar lipids in *Thermoplasma acidophilum* HO-62, *J. Bacteriol. 190*, 5404-5411.

75. Uda, I., Sugai, A., Itoh, Y. H., and Ito, T. (2001) Variation in molecular species of polar lipids from *Thermoplasma acidophilum* depends on growth temperature, *Lipids 36* 103-105.

76. Brassell, S. C., Wardroper, A. M. K., Thomson, I. D., Maxwell, J. R., and Eglinton, G. (1981) Specific acyclic isoprenoids as biological markers of methanogenic bacteria in marine sediments, *Nature 290*, 693-696.

77. Sturt, H. F., Summons, R. E., Smith, K., Elvert, M., and Hinrichs, K.-U. (2004) Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multi-stage mass spectrometry - new biomarkers for biogeochemistry and microbial ecology, *Rapid. Commun. Mass Spectrom. 18*, 617-628.

78. Brown, D. A., Venegas, B., Cooke, P. H., English, V., and Chong, P. L.-G. (2009) Bipolar tetraether archaeosomes exhibit unusual stability against autoclaving as studied by dynamic light scattering and electron microscopy, *Chem. Phys. Lipids 159*, 95-103.

79. Chong, P. L.-G. (2010) Archaebacterial bipolar tetraether lipids: physicochemical and membrane properties, *Chem. Phys. Lipids 163*, 253-265.

80. Komatsu, H., and Chong, P. L.-G. (1998) Low permeability of liposomal membranes composed of bipolar tetraether lipids from thermoacidophilic archaebacterium *Sulfolobus acidocaldarius, Biochemistry 37*, 107-115.

81. Sinninghe Damste', J. S., Rijpstra, W. I. C., Hopmans, E. C., Schouten, S., Balk, M., and Stams, A. J. M. (2007) Structural characterization of diabolic acid-based tetraester, tetraether, and mixed ether/ester, membrane-spanning lipids of bacteria from the order *Thermotogales*, *Arch. Microbiol. 188*, 629-641.

82. Boucher, Y., Kamekura, M., and Doolittle, W. F. (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea, *Mol. Microbiol. 52*, 515-527.

83. Houten, S. M., and Waterham, H. R. (2001) Nonorthologous gene displacement of phosphomevalonate kinase, *Mol. Genet. Metab. 72*, 273-276.

84. Herdendorf, T. J. and Miziorko, H. M. (2006) Phosphomevalonate kinase: functional investigation of the recombinant human enzyme, *Biochemistry 45*, 3235-3242.

85. Kaneda, K.,Kuzuyama, T., Takagi, M, Hakayawa, Y., and Seto, H. (2001) An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from *Streptomyces* sp. strain CL190, *Proc. Natl. Acad. Sci. USA 98*, 932-937.

86. Barkley, S. J., Cornish, R. M., and Poulter, C. D. (2004) Identification of an archaeal type II isopentenyl diphosphate isomerase in *Methanothermobacter thermautotrophicus*, *J. Bacteriol. 186*, 1811-1817.

87. Dutoit, R., de Ruyck, J., Durisotti, V., Legrain, C., Jacobs, E., and Wouters, J. (2008) Overexpression, physicochemical characterization, and modeling of a hyperthermophilic *Pyrococcus furiosis* type II IPP isomerase, *Proteins: Structure, Function and Bioinformatics 71 (4)*, 1699-1707.

88. Grochowski, L. L., Xu, H., and White, R. H. (2006) *Methanocaldoccus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate, *J. Bacteriol. 188*, 3192-3198.

89. Lombard, J., and Moreira, D. (2011) Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life, *Mol. Biol. Evol. 28 (1)*, 87-99.

90. Vinella, D., Brochier-Armanet, C., Loiseau, L., Talla, E., and Barras, F. (2009) Iron-sulfur (Fe/S) protein biogenesis: phylogenomic and genetic studies of A-type carriers, *PLoS Genet. 5*:e1000497.

91. Martin, W. and Russell, M. J. (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells, *Philos. Trans. R. Soc. Lond. B Biol. Sci. 358*, 59-83.

92. Mulkidjanian, A. Y., Galperin. M. Y., and Koonin, E. V. (2009) Co-evolution of primordial membranes and membrane proteins, *Trends Biochem. Sci. 34*, 206-215.

93. Koga, Y., Kyuragi, T., Nishihara, M., and Sone, N. (1998) Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent, *J. Mol. Evol. 46*, 54-63.

94. Pereto, J., Lopez-Garcia, P., and Moreira, D. (2004) Ancestral lipid biosynthesis and early membrane evolution, *Trends Biochem. Sci. 29*, 469-477.

95. Chen, M. and Poulter, C. D. (2010) Characterization of thermophilic archaeal isopentenyl phosphate kinases, *Biochemistry 49 (1)*, 207-217.

96. Pakhomova, S., Bartlett, S. G., Augustus, A., Kuzuyama, T., and Newcomer, M. E. (2008) Crystal structure of fosfomycin resistance kinase FomA from Streptomyces wedmorensis, *J. Biol. Chem. 283*, 28518-28526.

97. Pakhomova, S., Bartlett, S. G., Doerner, P. A., and Newcomer, M. E. (2011) Structural and biochemical insights into the mechanism of fosfomycin resistance kinase FomA, *Biochemistry 50 (32)*, 6909-6919.

98. Ramon-Maiques, S., Marina, A.m Gil-Ortiz, F., Fita, I., and Rubio, V. (2002) Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase family, during catalysis, *Structure 10*, 329-342.

99. Marco-Marin, C., Gil-Ortiz, F.,and Rubio, V. (2005) The crystal structure of Pyrococcus furiosis UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis, *J. Mol. Biol. 352*, 438-454.

100. Pakhomova, S., Bartlett, S. G., Doerner, P. A., and Newcomer, M. A. (2011) Structural and biochemical insights into the mechanism of fosfomycin phosphorylation by fosfomycin resistance kinase FomA, *Biochemistry 50 (32)*, 6909-6919.

CHAPTER 2

X-RAY STRUCTURES OF ISOPENTENYL PHOSPHATE KINASE

Introduction

Kinases or phosphotransferases are enzymes that catalyze the transfer of a phosphate group from a high-energy nucleotide donor such as ATP to a substrate. More than 500 kinases are known in humans, acting on small molecules such as lipids, carbohydrates, amino acids, and nucleotides to prime them for metabolic pathways or as part of signaling processes. Protein kinases constitute one of the largest groups of this type of enzymes, making up 2% of the human genome and modifying more than 30% of all human proteins *(1)*. Phosphorylation occurs most commonly on the hydroxyl groups of serine and threonine, and less commonly on tyrosine and histidine, resulting in a change in enzyme activity, cellular location or association with other proteins. A malfunction of any of these regulatory machineries often leads to various diseases and disorders, including cancer.

A typical protein kinase contains a conserved core consisting of two lobes, a smaller C-terminal lobe and a larger N-terminal lobe that together form a cleft that accommodates Mg^{2+}- or Mn^{2+}-bound ATP *(2,3)*. A glycine-rich loop (G-loop) covers the β- and γ-phosphates of ATP and performs a critical role in phosphoryl transfer and in

ATP/ADP exchange during the catalytic cycle. The more rigid and primarily helical C-terminal lobe serves as a docking site for substrate peptides or proteins. The N-terminal part of the substrate peptide lies in a groove between the αD and αF helices on one side and the αG helix on the other (Figure 2.1a). At the C-terminus of the substrate peptide, the residue right after the phosphorylated side chain is buried in a pocket formed by the P+1 loop, found at the C-terminus of an extended activation segment (Figure 2.1b). This segment is the most important regulatory element in protein kinases as it controls both substrate binding and catalytic efficiency. In the middle of this activation segment is the activation loop (also called the phosphorylation lip) that contains a serine, threonine or tyrosine residue that can be autophosphorylated or phosphorylated by other protein kinases. The APE motif at the C-terminus of the activation segment anchors the P+1 loop to the αF helix. This motif is conserved in almost all kinases and contains invariant Ala and Pro residues that bind to a conserved Trp residue in the αF helix *(5-8)*.

The proper positioning of protein kinase elements ensures that ATP and the substrate peptide are precisely positioned for catalysis, especially since phosphoryl transfer is very sensitive to the distance between the reactants. This is achieved by the presence of two noncontiguous hydrophobic spines, called the C-spine (catalytic spine) and the R-spine (regulatory spine) that define the internal architecture and conformation of all protein kinases and provide a firm yet flexible scaffold for these elements (Figure 2.2a and b). The C-spine starts at the C-terminus of the αF helix, traverses both lobes of the protein kinase molecule through a stretch of hydrophobic residues, and controls catalysis by localizing the ATP in its binding cleft. The R-spine, on the other hand, starts at a conserved Asp residue in the αF helix that hydrogen bonds with the main chain of

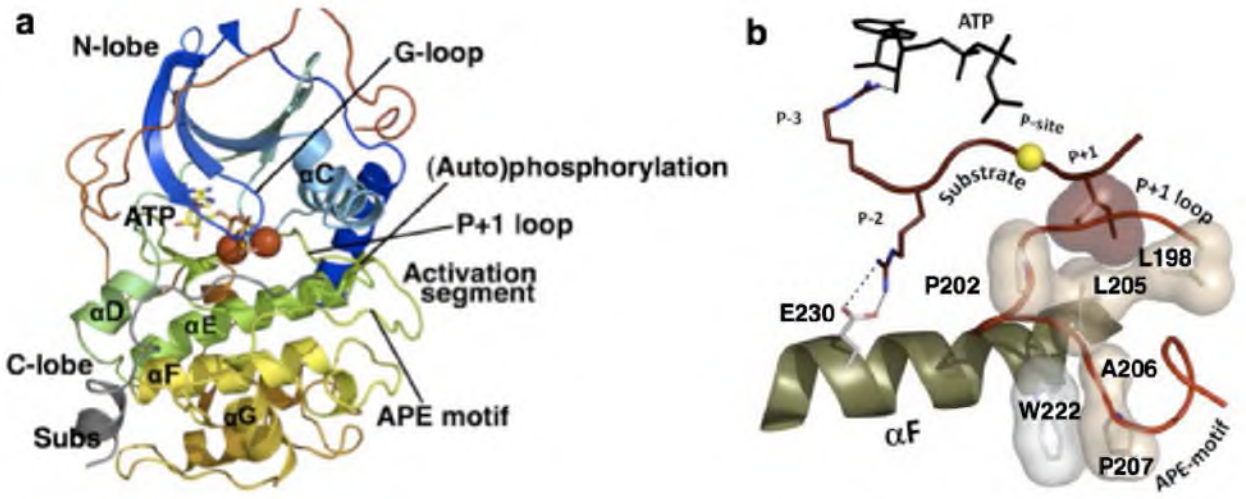


Figure 2.1. General architecture of a protein kinase, using PDB structure 1ATP. (a) The N- and C-terminal lobes of protein kinase form a cleft for binding ATP. The divalent metals Mg^{2+} and Mn^{2+} are shown as red spheres. The glycine-rich loop (G-loop) forms a lid over the β- and γ-phosphates of ATP and plays a role in phosphoryl transfer. The substrate peptide (grey) is located in the primarily helical C-terminal lobe, where its N-terminus is flanked by the αD and αF helix on one side and the αG helix on the other. The activation segment is the most important regulatory element in protein kinases and consists of the P+1 loop, the activation loop (phosphorylation lip) and the APE motif at the C-terminus. (b). The substrate peptide is positioned next to the αF helix by conserved hydrophobic residues in the activation segment. The P+1 loop forms a hydrophobic pocket anchored to Trp222 in the αF helix via the APE motif. The phosphorylation site on the substrate is shown as a yellow sphere. The P-2 arginine residue anchors the peptide to the αF helix by hydrogen bonding with Glu230.

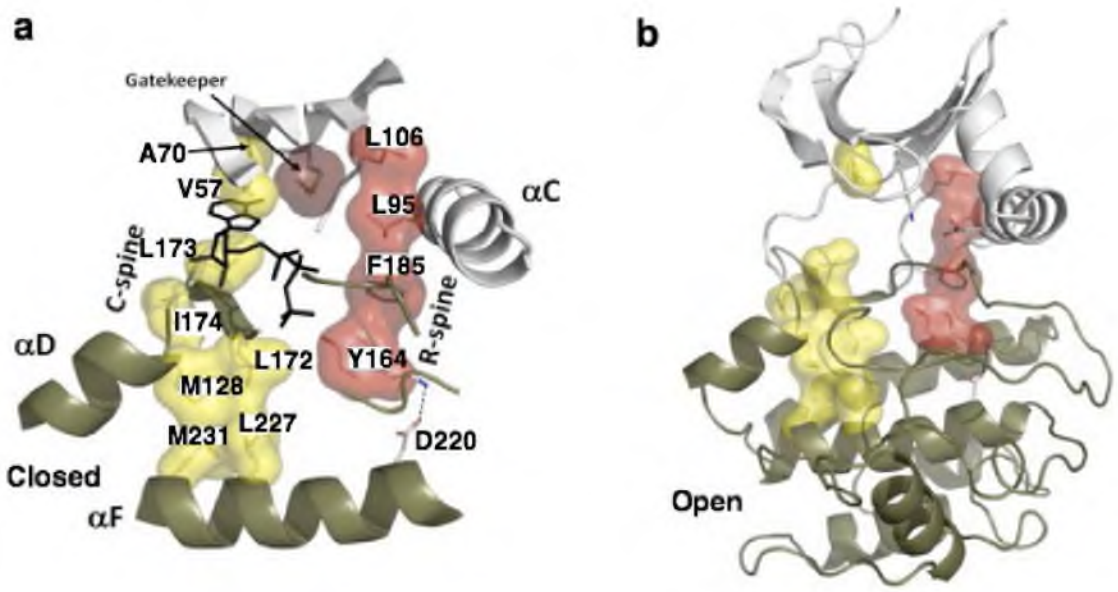


Figure 2.2. Two hydrophobic spines anchored to the αF helix that define the internal architecture of all protein kinases. Both spines span the N- and C-termini of the molecule through a series of side chains from non-contiguous secondary structures in the enzyme. (a) The catalytic spine (C-spine) is shown as a yellow surface. The R-spine is colored red. The gatekeeper residue (dark red) is located between the two spines and is spatially conserved. (b). Protein kinase in an open conformation. Efficient catalysis depends on the "breathing" of the protein kinase, during which both spines remain intact.

His/Tyr from the HRD motif, introducing a hydrophobic backbone of bulky residues found in the activation segment and the αC helix. Formation of the R-spine depends on phosphorylation of the activation segment, and all active protein kinases possess an R-spine while inactive ones do not. Interestingly, this spine model has allowed the identification of residues that occupy spatially conserved sites. One such example is a methionine "gatekeeper" residue, called as such because it was used to regulate the binding of unnatural ATP analogues (Figure 2.2a). While the gatekeeper residue is not conserved in all kinases and its side chain in many cases can be removed without loss of kinase activity, its substitution can lock the R-spine in its active, closed form. For instance, mutation of the Thr gatekeeper residue in Src and Abl kinases to bulkier side chains such as methionine and isoleucine led to a constitutively active oncogenic kinase because of a permanently locked R-spine *(9-15)*.

A smaller subset of the kinase family consists of enzymes that can phosphorylate small molecules with acyl groups, forming an acylphosphate. To date, four different "molecular machines" or structural devices are known to synthesize an acylphosphate product, having as paradigms NAGK, acetate kinase, phosphoglycerate kinase and biotin carboxylase, where each differs in architecture, domain folding and nucleotide binding (Figure 2.3). However, common among all enzymes in this subset are the short-chain carboxylic acid containing substrates, which are thought to have been among the most common organic molecules in prebiotic and early biotic environments. Acetic acid, for instance, has the interesting property of being capable of diffusing freely across membranes, being retained as a biosynthetic precursor, and being activated for biosynthetic reactions *(16)*.

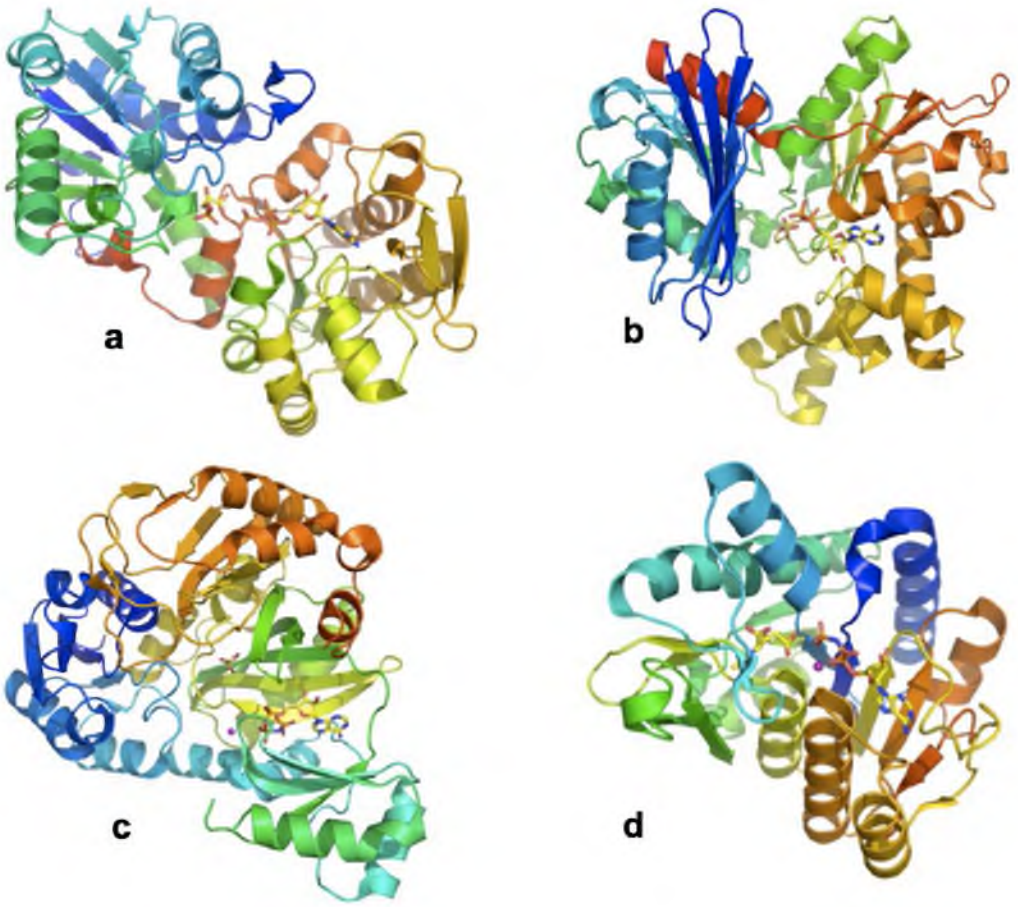


Figure 2.3. Four different kinds of molecular devices for forming acylphosphate bonds are known and are exemplified by the following enzymes: phosphoglycerate kinase (2X15) (a), acetate kinase (1G99) (b), the biotin carboxylase subunit of acetyl coA carboxylase (1BNC) (c), and N-acetyl-L-glutamate kinase (1GS5) (d). Bound nucleotides are shown in stick representation.

Phosphoglycerate kinase (PGK) is required for ATP synthesis in the glycolytic pathways of aerobes and anaerobes as well as for carbon fixation in plants. PGKs are monomeric enzymes with molecular masses of around 45 kilodaltons, with comparable amino acid compositions and similar catalytic properties. The most striking feature of PGK is that its single polypeptide chain is organized into two widely separated domains of almost equal size that make large-scale hinge-bending conformational changes and domain closure broadly similar to that reported for hexokinase (Figure 2.3a). The N-terminal domain of the enzyme binds 3-phosphoglycerate while the C-terminal domain binds ATP. Each domain is composed of a central β sheet of six parallel strands surrounded by helices, although there are no significant similarities in their sequences and their detailed structures are different. As in other glycolytic enzymes, the overall tertiary structure is formed by the repetition of β-α units, which is repeated 12 times in PGK *(17)*.

Acetate kinase (AK) represents the second type of acylphosphate-forming device. AK converts acetyl phosphate and ADP to acetate and ATP in the production of methane, and converts acetate to acetyl phosphate in the conversion of acetate to methane and carbon dioxide. The N-terminal domain of AK consists of an eight-stranded β sheet and eight α helices, while the C-terminal domain consists of a seven-stranded β sheet, eleven helices and an additional small two-stranded β sheet (Figure 2.3b). The ATP binding site is located in the cleft between the two domains. The AK fold consists of a core identical to that of the glycerol kinase/hexokinase/actin/Hsc70 superfamily (also called the ASKHA superfamily). This core is composed of a duplicated βββαβαβα secondary

structure with insertions of subdomains between particular elements of the β sheet. The residues that bind the magnesium ion are also conserved among ASKHA enzymes *(16)*.

Biotin carboxylase (BC) is a subunit of acetyl CoA carboxylase, which catalyzes the first committed step and one of the regulated steps in the biosynthesis of long-chain fatty acids. The subunit biotin carboxylase, on the other hand, catalyzes the biotin-dependent carboxylation of acetyl CoA to form malonyl CoA in a two-step reaction mechanism. Here, the reaction using bicarbonate and ATP forms a carboxyphosphate intermediate that collapses immediately into phosphate and CO_2 with subsequent attack of biotin on the latter. The BC monomer is asymmetric and consists of three structural motifs (Figure 2.3c). The N-terminal domain adopts a dinucleotide binding motif consisting of five strands of parallel β-pleated sheet flanked on either side by a total of four α helices. This architecture is similar to that of NAD^+-dependent dehydrogenases and UDP-galactose 4-epimerase and was unexpected since BC is ATP-dependent. In addition to the β strands and α helices, the N-terminal domain contains six type I turns. The B domain follows the N-terminal domain and projects outward of the main body of the subunit. It contains two α helices, three strands of antiparallel β sheets and flexible interconnecting loops rich in glycine residues. Finally, the C-terminal domain is an eight-stranded antiparallel β sheet with two β strands being disrupted by β bulges. It also contains a smaller three-stranded antiparallel β sheet, seven α helices and seven reverse turns. The C-terminal domain cradles the N-terminal nucleotide-binding domain, forming a distinct cleft in the biotin carboxylase structure *(18)*.

N-acetylglutamate kinase (NAGK) exemplifies the fourth kind of molecular device for making acylphosphates. It is part of a pathway in plants and microorganisms

that converts N-acetylated glutamate (NAG) to L-ornithine, the precursor of arginine. In animals, L-ornithine is produced by a similar pathway that utilizes nonacetylated glutamate. This differential trait and the fact that NAGK catalyzes an obligatory, and in many cases, controlling step in organisms that synthesize arginine via NAG, make it an appropriate target for selective inhibition of arginine biosynthesis.

A NAGK subunit consists of the classical αβα sandwich fold that is formed by a core β sheet of eight parallel strands caught between two layers of three and four α helices (Figure 2.3d). The core β sheet spans across the dimer interface to form a continuous 16-stranded molecular β sheet. The N-terminal lobe of each subunit serves as the NAG binding site, while the ATP binding site is found in the C-terminal lobe. Each subunit contains one catalytic center, and while dimer formation is elaborate and formed by many interactions involving six highly conserved residues, it is not known to facilitate substrate channeling. Rather, its role is purely structural, providing a scaffold for anchoring the complex loop system that forms the substrate-binding site in each subunit *(19)*.

The classical αβα sandwich fold is seen in other enzymes important in the biosynthesis of other amino acids, nucleotides and important metabolites (Figure 2.4). These enzymes form the amino acid kinase family (AAK) and include carbamate kinase (CK) *(20)*, the N-terminal domain of aspartokinase (AK) *(21)*, glutamate-5-kinase (G5K) *(22)*, and uridylate monophosphate kinase (UMPK) *(23)*. The recently determined structure of the bacterial resistance enzyme FomA in *S. wedmorensis* also has this characteristic fold. IPK from THA, MTH and MJ are expected to possess the αβα sandwich fold based on sequence alignments with NAGK (Figure 2.5). This chapter

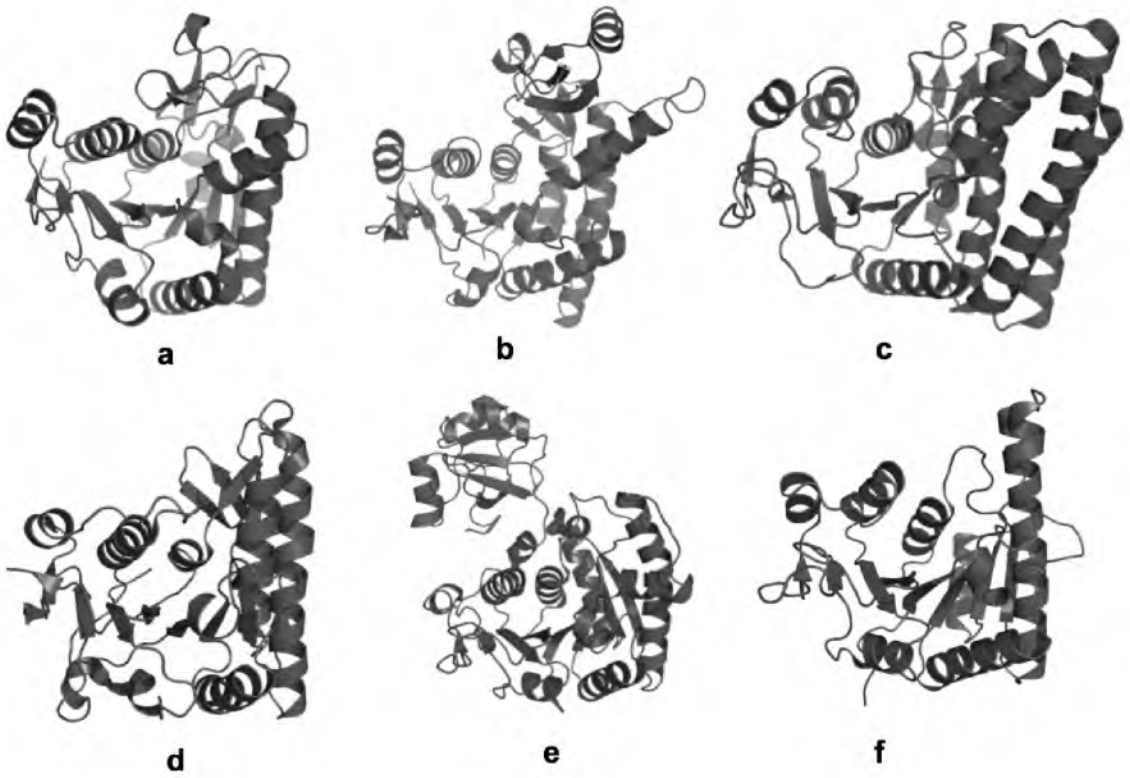


Figure 2.4. The αβα sandwich fold is characteristic of enzymes in the amino acid kinase (AAK) family. (a) N-acetyl-L-glutamate kinase (NAGK) (1GS5). (b) Carbamate kinase (CK) (2WE4). (c) The N-terminal domain of aspartokinase (AK) (2JOX). (d) L-glutamate-5-kinase (G5K) (2J5T). (e) Uridine monophosphate kinase (UMPK) (2BMU). (f) Fosfomycin kinase (FomA) (3D41).

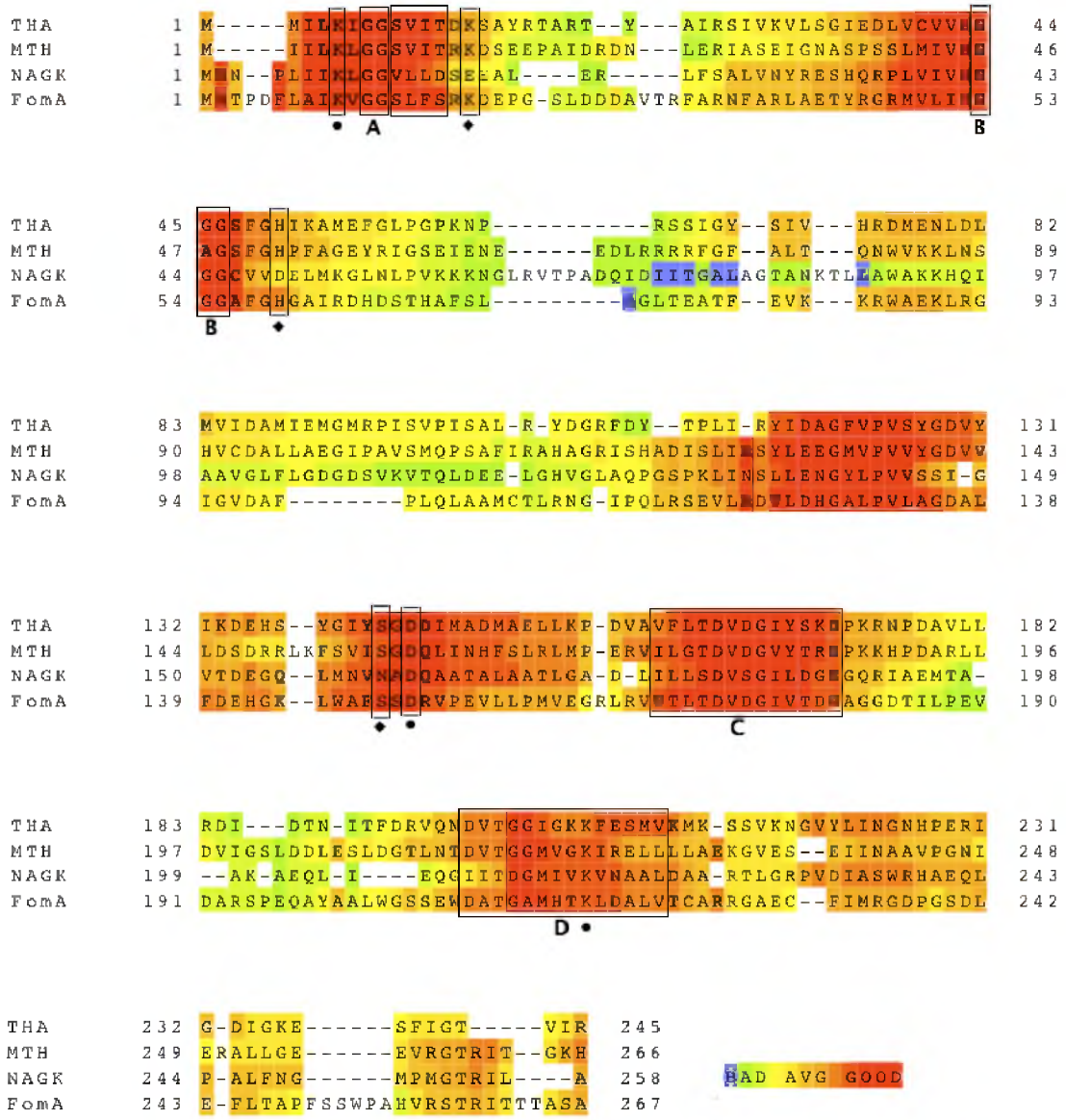


Figure 2.5. Sequence alignment of four enzymes of the amino acid kinase family shows conservation of catalytic and binding residues. The degree of conservation of residues is color-coded, with red signifying highest conservation among the compared sequences. Residues marked with (•) are known in *E. coli* NAGK to perform key organizing and transition state stabilizing roles in the phosphoryl transfer reaction, and are also found in THA and MTH IPKs and in FomA. Residues marked with (♦) are found only in IPKs and in FomA. Glycine-rich loops that facilitate phosphoryl transfer are labeled A and B, while ATP binding regions are labeled C and D.

presents the first structures of THA and MTH IPKs, as apoenzyme in the case THA IPK and in ternary complex with substrates and products for both MTH and THA IPKs. The structures that will be presented confirm the structural prediction as well as elucidate the key similarities and differences between archaeal IPKs and other enzymes in the AAK family.

Experimental Procedures

Gene amplification and cloning. The THA IPK gene was amplified from genomic DNA (ATCC: 25905D) by PCR using *Pfu*Ultra II DNA polymerase (Stratagene) and the forward and reverse primers 5'–CACCATGATGATACTGAAGATAGG–3' and 5'–TCATCTTATCACCGTACCTATGAATG–3', respectively. The forward primer contains the sequence CACC, which anneals with the overhang sequence of the topoisomerase-charged cloning vector, pET151/D-TOPO (Invitrogen). The double stranded PCR product was directionally cloned into the pET151/D-TOPO vector to append an N-terminal hexahistidine tag, the V5 epitope, and a TEV protease cleavage sequence upstream of the start codon. The resulting construct was used to transform *E. coli* DH5α cells, which were then grown overnight in LB plates containing ampicillin. Single colonies were picked for purification of plasmids containing the THA IPK gene. Plasmids containing the correct insert were used to transform *E. coli* BL21 (DE3)-CodonPlus-RIL cells (Stratagene) for repeatable and high-yield expression. The amplification and cloning of the MTH IPK gene is described in reference *(24)*.

Protein expression and purification. Selenomethionine-substituted THA IPK was produced in *E. coli* BL21 (DE3)-CodonPlus-RIL (Stratagene) using the autoinduction

method *(25)*. A starter culture was grown overnight in PA-0.5G media containing chloramphenicol and ampicillin, and then diluted 1000-fold in PASM-5052 media containing 25 mg mL^{-1} selenomethionine, 1x trace metals mix, 100 μM vitamin B_{12} and 17 amino acids (each at 10 mg mL^{-1}) excluding cysteine, tyrosine and methionine. The culture was incubated at 37 °C for 8 h followed by incubation at 23 °C until OD_{600} > 8.0. Cells were harvested by centrifugation, resuspended and lysed in binding buffer containing 50 mM Na_2HPO_4, pH 8.0, 150 mM NaCl, 10 mM imidazole and 1 mM β-mercaptoethanol. DNAse (1 μg mL^{-1}) and lysozyme (1 mg mL^{-1}) were added to the cell suspension and incubated for 30 min. The cell suspension was sonicated for three min, and cell debris was removed by centrifugation (20,000 x g for 25 min at 4 ^{o}C). The supernatant was heat-treated for 15 min at 50 °C before another centrifugation step. The clarified supernatant was incubated at 4 °C with Ni-NTA resin (Qiagen) preequilibrated with binding buffer. Bound proteins were eluted with buffer containing 50 mM Na_2HPO_4, pH 8.0, 150 mM NaCl, 250 mM imidazole and 1 mM β-mercaptoethanol. Eluted proteins were pooled and the hexahistidine tag and V5 epitope were removed by adding recombinant TEV protease at 1 mM with dialysis against a buffer solution containing 50mM Tris-HCl, 150mM NaCl and 1mM DTT at 4 °C for 24 h. The dialysate was then subjected to another round of dialysis using the same buffer without DTT for another 4 h. The resulting dialysate was then applied to Ni-NTA and the flow through containing the cleaved THA IPK was subjected to a final purification step by gel filtration using Superdex 200 (GE Healthcare) equilibrated with 50 mM HEPES pH 7.5, 150 mM NaCl, 1mM β-mercaptoethanol. Using the extinction coefficient 17,880 M^{-1} cm^{-1} at 280 nm calculated using the ProtParam tool in the ExPASy Proteomics server *(26)*, THA IPK

concentration was estimated and protein was concentrated to 0.30 mM for subsequent crystallization. Selenomethione incorporation of cleaved THA IPK was confirmed by mass spectrometry. Native THA IPK was obtained by growing transformed BL21 (DE3)-CodonPlus-RIL cells (Stratagene) in ZY-5052 autoinduction media and purified using the same procedure for THA IPK.

The expression and purification of MTH IPK is described in *(24)*. For crystallization, the purified protein was concentrated to 5 mg mL^{-1} in a stabilization buffer of 25 mM HEPES, pH 7.5, 75 mM NaCl, 0.5 mM β-mercaptoethanol.

Crystallization. All crystals of THA IPK were grown by the sitting drop vapor diffusion method at 21 °C using a mixture consisting of 0.30 mM protein, 1.5 mM $MgCl_2$, 1.5 mM IP and 1.5 mM ATP that was prepared and incubated for at least two h before crystallization. In all cases the crystals grew to maximum dimensions in about one week in drops of a 1:1 mixture of protein and reservoir solutions. Selenomethionine-substituted THA IPK crystals complexed with IP and ADP were grown using a solution of 0.1 M sodium citrate, pH 5.0, and 15% PEG 6000. Crystals grew to maximum dimensions in about one week. The IPK•IP•ATP crystal was grown using native protein and a reservoir solution of 0.1M MIB buffer (2:3:3 molar ratio of sodium malonate, imidazole and boric acid), pH 7.0 and 25% PEG 1500. The IPK•IPP•ADP crystal was obtained using selenomethionine-substituted protein and a reservoir solution of 0.1M PCB buffer (2:1:2 molar ratio of sodium propionate, sodium cacodylate and Bis-Tris propane) pH 6.0 containing 25% PEG 1500. In this case the substrate included with the crystallized protein was converted and preferentially bound as product under the crystallization conditions. Crystals of MTH IPK were grown by mixing 2 μL protein (5

mg mL^{-1} MTH IPK, 5 mM $MgCl_2$, 1 mM ATP) with a precipitant solution containing 12% PEG 6000 and 2M NaCl.

Crystal structure determination. Prior to data collection, all crystals were cryoprotected by immersion for 30 s in a modified reservoir solution that contained 30% glycerol and cooled by plunging into liquid nitrogen. Diffraction data were collected at 100 K on a rotating anode source or at beamline 11-1 of the SSRL and processed using HKL2000 *(27)*. The THA IPK structure in complex with IP and ADP was determined by single wavelength anomalous diffraction (SAD) using PHENIX *(28)*. Model building was done using COOT *(29)*. Refinement used REFMAC5 *(30)* with the final refinement cycles performed using PHENIX. This refined model was used to determine all of the other structures by molecular replacement using PHASER *(31)* in the CCP4 suite *(32)*.

The DaliLite server *(33)* was used to superimpose the Cα atoms of two protein structures and measure the RMSD between the two. Molprobity *(34)* was used for structure validation. PyMol *(35)* and the command "Ray" were used to create images of the protein structures.

Metal ion requirement. The metal ion dependence of IP kinase was determined by the radioactivity assay using IP and [^{32}P] ATP (American Radiolabeled Chemicals) as substrates *(24)*. Reaction mixtures contained solutions prepared using 18 MΩ deionized water, including 100 mM HEPES, pH 7.5, 10 mM β-mercaptoethanol, 0.1% (w/v) BSA, 400 μM IP, 400 μM [^{32}P] ATP and 10 mM of $MgCl_2$, $MnCl_2$, $ZnCl_2$, $CuCl_2$, $NiCl_2$, $CdCl_2$ or $CoCl_2$ (atomic absorption grade, Alfa Aesar). The reactions were initiated by adding purified THA IPK to a final volume of 50 μL followed by incubation at 37 °C for 10 min. Reactions were quenched by adding 113 μL of methanol/750 mM EDTA (100:13, v/v).

Samples (6 μL) of the quenched reaction mixtures were spotted on silica TLC plates and developed with $CHCl_3$/pyridine/formic acid/H_2O (30:70:16:10, v/v). The TLC plates were imaged for 24 h using a storage phosphor imager autoradiography cassette and visualized by a Typhoon 8600 variable mode imager.

Results

Gene amplification and cloning. Directional TOPO® Cloning (Invitrogen) was used to incorporate the THA IPK gene to the pET151/D-TOPO vector. The previous kinetic studies by Chen and Poulter relied on pET28b-based constructs for both THA and MTH IPK. While pET28b-MTH IPK yielded well-diffracting crystals that resulted in a native diffraction data set to 2.15 Å, pET28b-THA IPK yielded highly mosaic crystals with overlapping diffraction spots that made data reduction inaccurate. On the other hand, the pET151/D-TOPO-THA IPK construct proved to be desirable, eventually leading to quality crystals that diffracted to as low as 1.90 Å.

Directional TOPO® Cloning is based on the ability of topoisomerase I from *Vaccinia* virus to bind duplex DNA at specific sites and cleave the phosphodiester backbone after 5'-CCCTT in one strand. The energy from the broken phosphodiester backbone is conserved by the formation of a covalent bond between the 3' phosphate of the cleaved strand and Tyr274 of topoisomerase I. The enzyme-DNA phosphotyrosyl bond can be attacked by the 5' hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase. Directional joining of the double-stranded THA IPK gene to pET151/D-TOPO is made possible by adding a 5'-CACC sequence to the insert, which is complementary to the 3'-GTGG overhang in the cloning vector. The

overhang in the pET151/D-TOPO vector invades the 5'-end of the THA IPK insert and stabilizes it in the correct orientation. The insert was then cloned into the pET151/D-TOPO vector and the duplex sequence was verified to be correct.

Protein expression and purification. Constructs of either MTH or THA IPK based on the pET28b vector always resulted in overexpressed proteins with greater than 95% purity after SDS PAGE, and size exclusion chromatography consistently increased the purity to almost 99%. Due to the inability of the pET28b-THA IPK construct to yield quality crystals, the pET151/D-TOPO-THA IPK construct was created. Unfortunately, even after optimization of purification steps that included heat treatment, overexpressed THA IPK coeluted with many cell impurities during SDS PAGE (Figure 2.6a). Cleavage of the histidine tag and the V5 epitope by TEV protease followed by another metal ion chromatography step diminished the amount of impurities, leading to an enriched THA IPK with greater than 95% purity (Figure 2.6b). Size exclusion chromatography using Superdex 200 removed the remaining impurities (Figure 2.6c) yielding ultrapure THA IPK suitable for crystallization trials. Like MTH IPK, THA IPK is dimeric in solution, as shown by size exclusion chromatography using a column calibrated with protein standards. For cells grown in the presence of selenomethionine, mass spectrometry confirmed the substitution of all 12 methione residues in each THA IPK subunit with selenomethionine, making this derivative suitable for structure determination via anomalous scattering experiments.

Crystallization and structure determination. We have determined structures of IP kinases from two archaeal species, THA and MTH, with crystals obtained under a variety of conditions. All crystals were grown in the presence of IP and ATP. Although

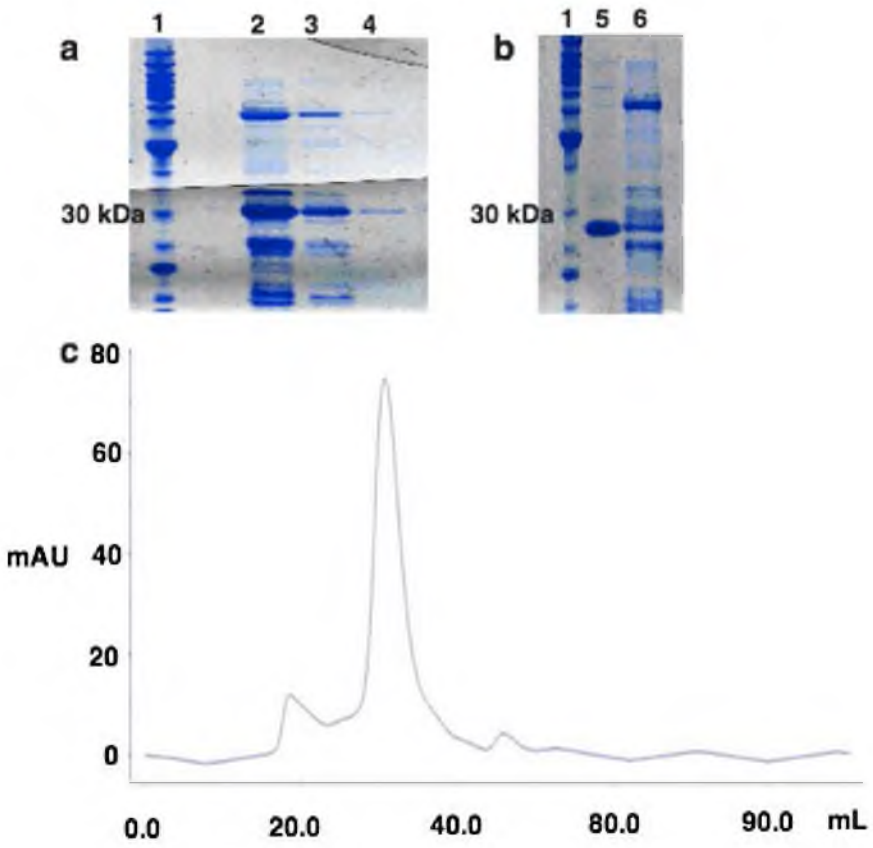


Figure 2.6. Purification of THA IPK prior to crystallization requires three steps. (a) SDS PAGE after first Ni affinity chromatography step shows overproduction of THA IPK. (1) Molecular weight markers, (2-4) fractions eluted with 250 mM imidazole. (b) Digestion of the histidine tag by TEV protease leads to further purification of THA IPK. (5) Flow through containing tagless THA IPK, (6) fraction eluted with 250 mM imidazole after TEV cleavage. (c) Size exclusion chromatogram showing a large peak for the THA IPK dimer.

structures of homologous proteins have been determined and are available in the Protein Data Bank (PDB), none of these structures proved suitable for molecular replacement even after the removal of nonconserved motifs. Isomorphous replacement using Hg, Sm, and Gd crystal soaks also did not yield sufficient phasing power or correct phases for structure determination. Thus, the THA IPK structure was determined by single wavelength anomalous diffraction (SAD) using crystals of selenomethionine-substituted protein grown in 0.1 M citrate buffer, pH 5.0 (Figure 2.7a). Data were collected from the Stanford Synchrotron Radiation Lightsource (SSRL) Beam Line 11-1, and a sample diffraction image from a selenomethionine derivative crystal with resolution up to 1.90 Å is shown in Figure 2.7b. Data for multiwavelength anomalous diffraction (MAD) phasing were collected after performing a fluorescence scan on a selenomethionine derivative crystal to identity the Se peak (0.97644 Å) and inflection (0.97934 Å) wavelengths, as well as a third "remote" wavelength (1.5418 Å), at which to collect anomalous data sets (Figure 2.7c). Despite the availability of three anomalous data sets, MAD phasing did not yield an interpretable electron density map. SAD phasing using the Se peak data set located a total of 45 selenium sites that were used to calculate initial phases to 1.90 Å. The phase estimates were improved by density modification to generate a readily interpretable electron density map where almost 70% of the THA IPK structure was autobuilt by PHENIX (Figure 2.8a). The rest of the protein structure was built manually and wrong connectivities were corrected in COOT. The structure of THA IPK with labeled secondary structures is shown in Figure 2.8b *(36)*.

The crystal asymmetric unit of the SAD structure contained two molecules of IPK, each of which contained IP and ADP in the active site (Figure 2.9). This

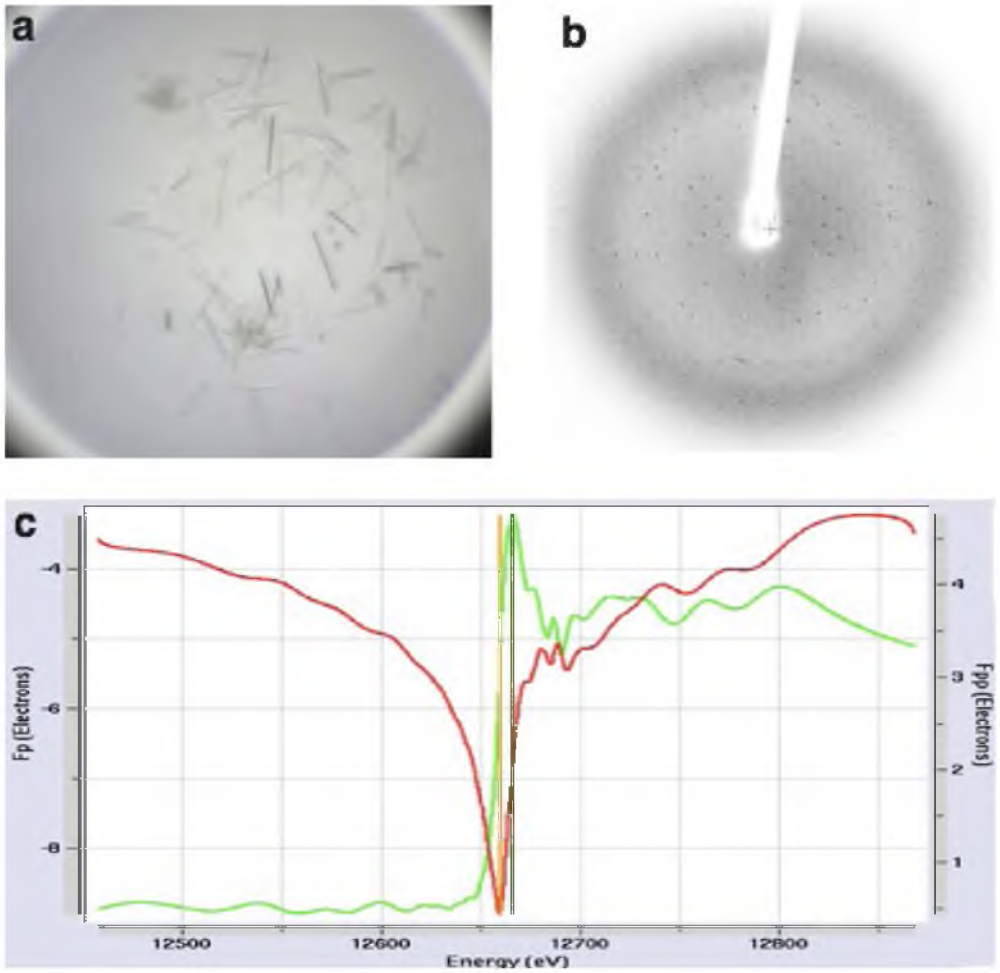


Figure 2.7. Multiwavelength anomalous diffraction data collection from crystals of selenomethionine-substituted THA IPK. (a) Rod-like crystals of THA IPK grown in 0.1 M citrate buffer, pH 5.0. (b) Sample image showing well-defined diffraction spots up to 2.0 Å. (d) Kramers-Kronig transformation of a fluorescence scan of a selenomethionine crystal used to identify the Se peak and inflection wavelengths, as well as a remote wavelength, at which to collect diffraction data.

combination of ligands presumably resulted from hydrolysis of ATP that occurred during sample preparation and crystallization and may constitute a dead end complex. This refined structure was used as a search model in subsequent molecular replacement solutions of other IPK crystal forms. Comparison of this IPK•IP•ADP dead end complex with other structures that were subsequently solved showed no significant differences in tertiary structure and conformation of important active site residues. Thus, in the following discussion of the structure of IPKs, reference will be made to more interesting and complete structures that contain the substrates IP and ATP, as well as the products IPP and ADP, unless otherwise noted.

THA IPK crystals grown in 0.1 M malonate/imidazole/borate (MIB) buffer, pH 7.0, contained two molecules of the enzyme in the asymmetric unit (Figure 2.10a). The two asymmetric unit molecules were almost identical and both contained IP and ATP in the active site, and will be called IPK•IP•ATP (Figure 2.10b and c). An F_o-F_c omit map contoured at 1σ confirms the identity of these ligands (Figure 2.10d). In addition, THA IPK crystals grown in propionate/cacodylate/Bis-Tris (PCB) buffer, pH 6.0 contained four molecules in the asymmetric unit (Figure 2.11a) where two molecules contained the products IPP and ADP (Figure 2.11b, c, and d) formed during crystallization, while the other two contained the substrates, IP and ATP (Figure 2.11e and f). An F_o-F_c omit map shows the proper assignment of the products IPP and ADP to the electron density (Figure 2.11b). The presence of substrate and product complexes in the asymmetric unit is reasonable given an equilibrium constant for the reaction, K_{eq} = 6.3, for uncomplexed substrates and products *(24)*. A structure for MTH IPK with two molecules in the asymmetric was also determined (Figure 2.12a). The asymmetric unit molecules

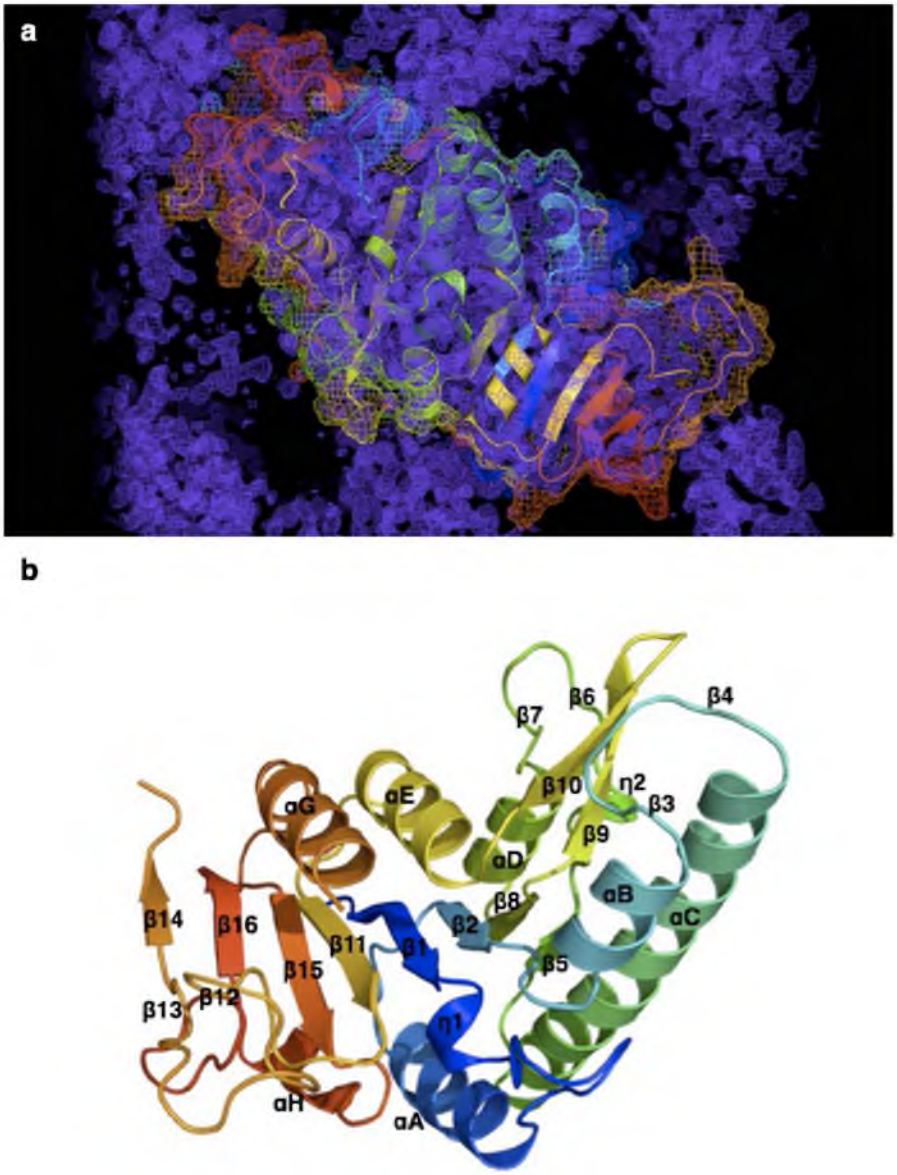


Figure 2.8. The tertiary structure of THA IPK contains the classical αβα sandwich fold found in other amino acid kinase enzymes. (a). $2F_o$-F_c electron density map showing the crystal lattice and a dimeric THA IPK built into the electron density. (b). Ribbon representation of THA IPK with labeled secondary structure, where all the elements found in the NAGK enzyme are also found in THA IPK.

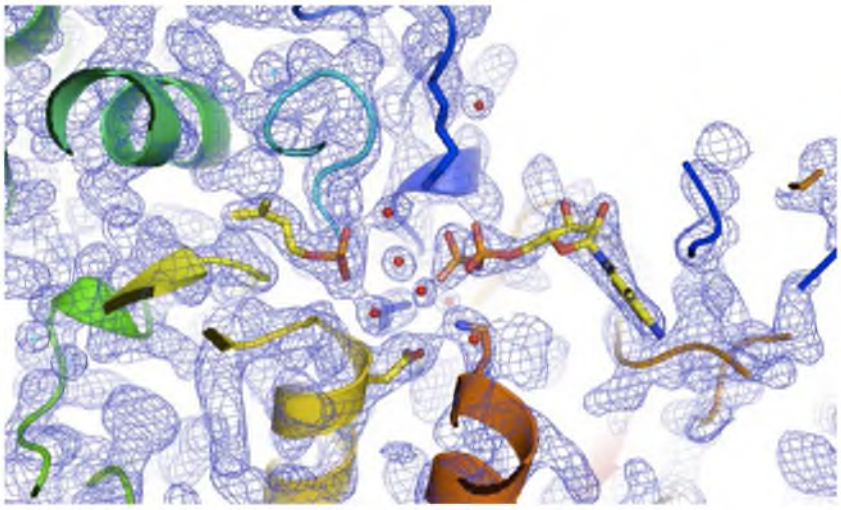


Figure 2.9. $2F_o$-F_c map of the dead-end complex showing bound substrate IP and product ADP. The structure was solved by single wavelength anomalous diffraction and was used as a model to solve other structures by molecular replacement. Water molecules in the cavity between IP and ADP are shown as red spheres.

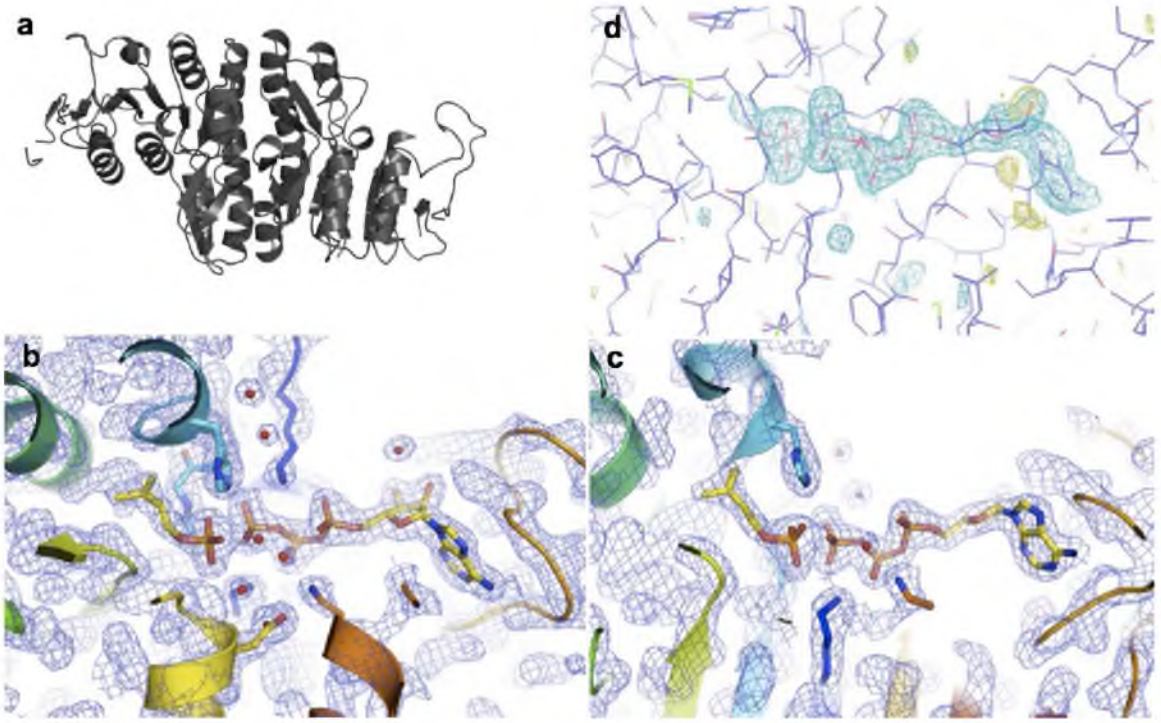


Figure 2.10. 2.0 Å crystal structure of the THA IPK substrate complex. (a) The crystal asymmetric unit contained two molecules of IPK that form a dimer. (b) $2F_o$-F_c map showing bound substrates IP and ATP, with active site water molecules shown as red spheres. (c) Alternate view of the active site showing the core β-sheet structure of IPK. (d) F_o-F_c ligand omit map showing electron density with modeled IP and ATP, indicating correct assignment of ligands.

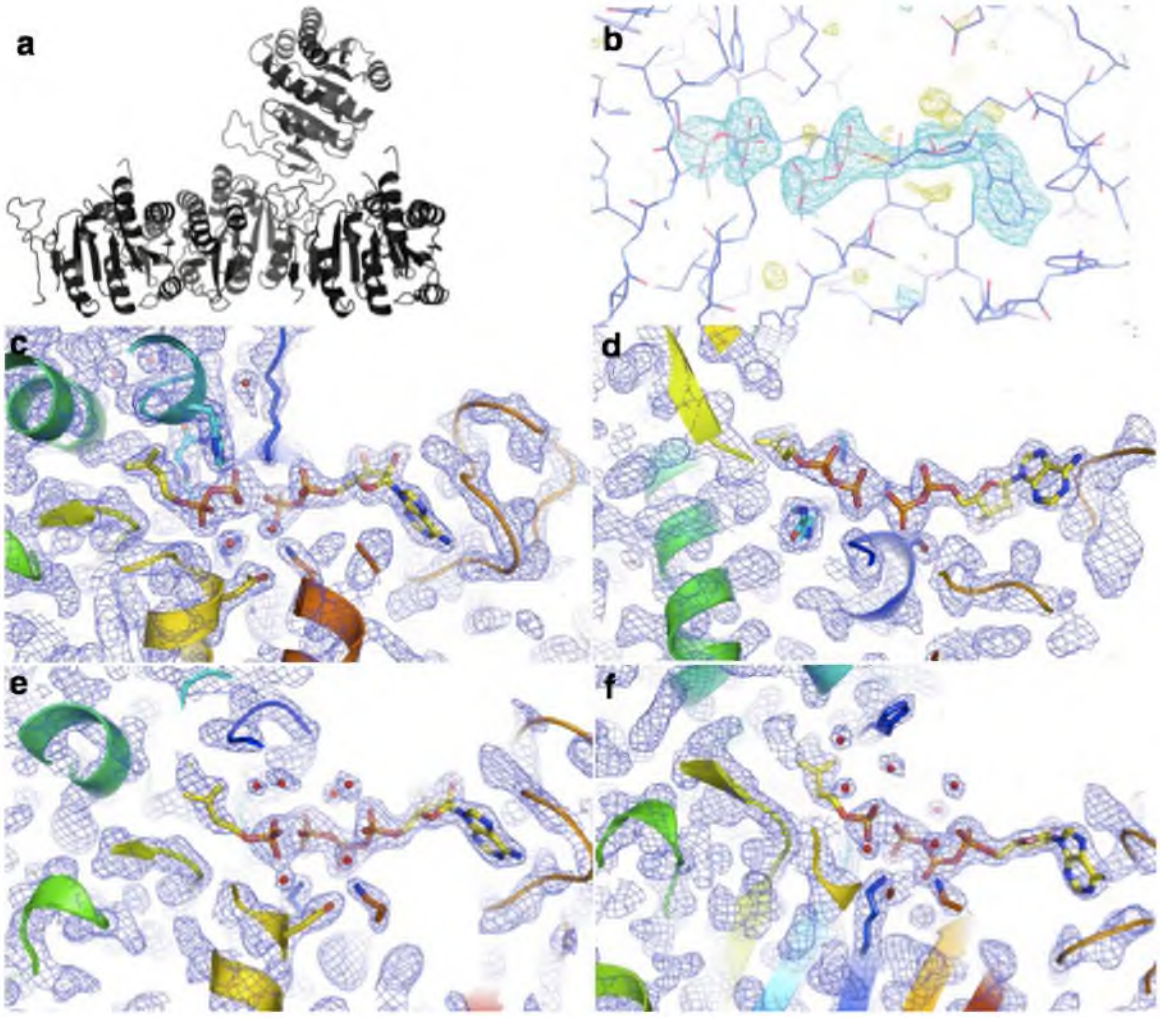


Figure 2.11. 1.99 Å crystal structure of the THA IPK product complex. (a) The crystal asymmetric unit contained four molecules. (b) F_o-F_c ligand omit map showing correct assignment of bound products, IPP and ADP to the electron density. (c) and (d) Two molecules in the crystal asymmetric unit contained bound products, IPP and ADP, shown above in two alternate views. (e) and (f) The other two molecules in the crystal asymmetric unit contained the substrates, IP and ATP. The presence of both substrates and products in the asymmetric unit was reasonable given the equilibrium constant between substrates and products, K_{eq} = 6.3. Water molecules are shown as red spheres.

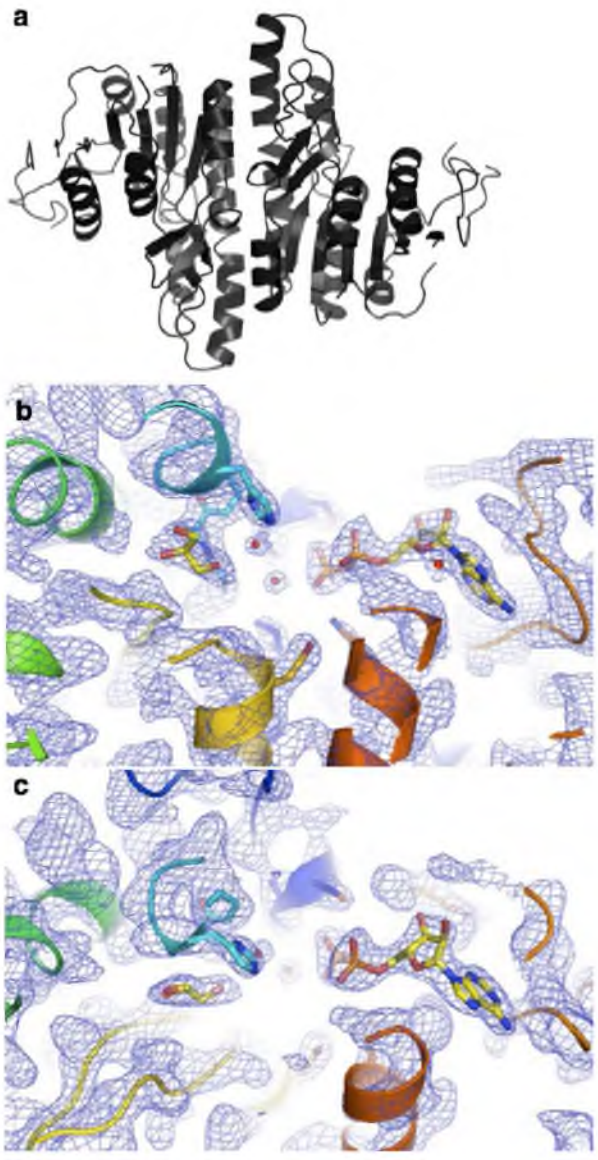


Figure 2.12. 2.15 Å crystal structures of MTH IPK. (a) The crystal asymmetric unit contained two molecules of MTH IPK that form a dimer. (b) and (c) $2F_o$-F_c electron density map showing bound product, ADP and a glycerol molecule that presumably displaced IPP during crystal mounting. Water molecules are shown as red spheres.

contained ADP and either glycerol or water molecules in the IP binding site (Figure 2.12b and c). Glycerol presumably displaced a bound IP molecule when the crystal was being prepared for cryogenic data collection. Finally, the structure of the THA IPK apoenzyme with six molecules in the asymmetric unit was determined from a crystal grown in 2.0 M $(NH_4)_2SO_4$ (Figure 2.13a). A representative subunit contains two sulfate molecules that are each positional mimics of the terminal phosphate of IP and the β phosphate of ATP, and molecules of water. The His50 residue was positioned in a different conformation than those in other structures (Figure 2.13b). All of the five structures were refined to R_{free} values of 23–27% against data to 2.0–2.7 Å resolution. All structures have excellent geometry, with more than 95% of the residues in favored Ramachandran regions. Crystallographic and refinement statistics for each structure are given in Table 2.1.

All of the crystallized proteins of THA IPK included N-terminal Phe(-1) and Thr(0) residues that were the remnants of a TEV protease recognition site. In some cases, one or both of these residues were visible in the electron density. Moreover, electron density for residues Asp(-3) and Pro(-2), also parts of the TEV protease recognition site, were found in the asymmetric unit of the crystal grown in PCB buffer. This asymmetric unit was interesting because of the combination of bound substrates and products in the molecules mentioned earlier, as well as the significant movement of loops in two molecules. Briefly, molecules A, B contained bound products, while molecules C and D contained bound substrates. In molecules C and D, the η1-αA loop was in a conformation different than those in molecules A and B (Figure 2.14a), which allowed these molecules to make crystal lattice contacts through Tyr17 (Figure 2.14b). This loop

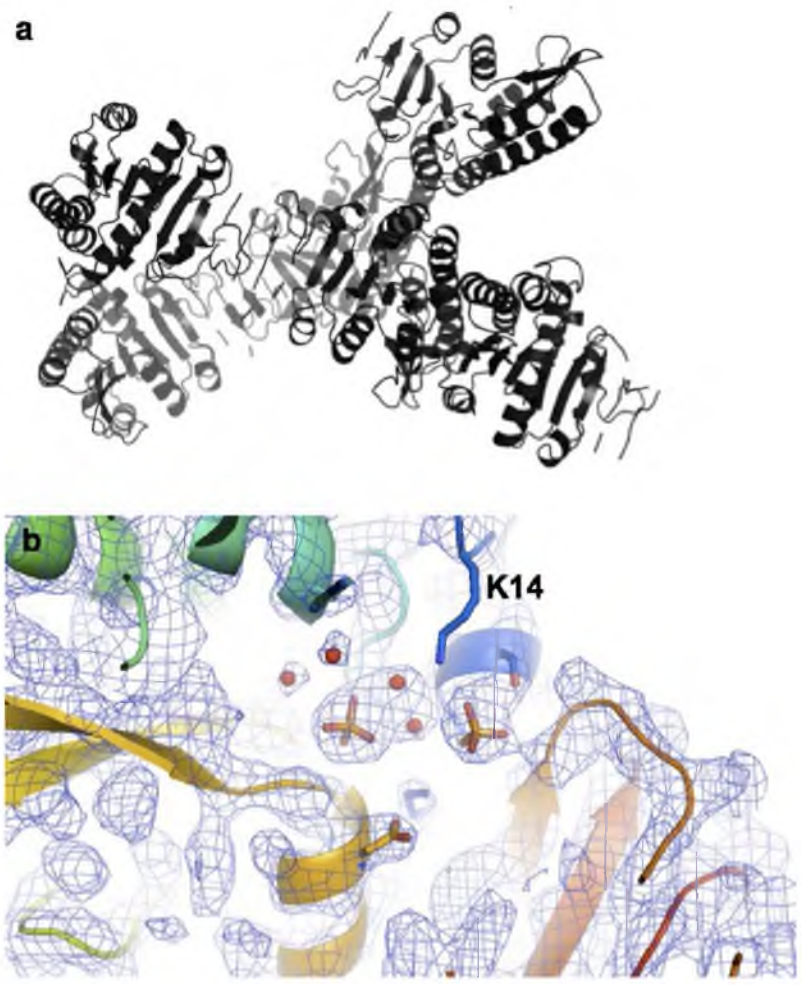


Figure 2.13. 2.7 Å crystal structure of the THA IPK apoenzyme. (a) Six THA IPK subunits formed the crystal asymmetric unit, containing water molecules (red spheres) and sulfate molecules (sticks). (b) $2F_o$-F_c electron density map showing bound sulfate molecules that mimic the positions of the IP phosphate group and the β-phosphate group of ATP. No clear electron density was observed for Lys14 presumably due to multiple conformations arising from the absence of ATP.

Table 2.1. Data collection and refinement statistics. Statistics for highest resolution shell are in parentheses. [a]$R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections. [b]$R_{factor} = \Sigma ||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

	IPK•IP•ATP	**IPK•IPP•ADP**	**MTH IPK**
Space Group	C2	C2	C2
Cell Dimensions	a = 124.31 b = 44.48 c = 91.76 β = 109.45	a = 187.10 b = 42.79 c = 134.90 β = 113.19	a = 119.67 b = 96.36 c = 72.43 β = 120.59
Molecules/ asu	2	4	2
Wavelength (Å)	1.54	0.97887	1.54
Resolution range (Å) (outer shell)	29.30 – 2.00 (2.07 – 2.00)	50.00 – 1.99 (2.06 – 1.99)	30.00 – 2.15 (2.18 – 2.15)
Unique reflections	31, 495	68, 827	38, 177
R_{sym}[a] (%)	7.5 (4.5)	6.20 (23.4)	9.1 (45.4)
Completeness	98.0 (95.2)	94.6 (80.9)	92.3 (66.2)
Resolution range	29.30 – 2.00	38.29 – 1.99	30.00 – 2.15
R_{factor}[b](work/test) (%)	17.6/23.0	19.2/23.4	20.7/ 26.0
RMSD bonds (Å)	0.018	0.019	0.022
RMSD angles (deg)	1.7	1.9	2.0
Most favored (%)	98.70	97.81	94.02
Allowed (%)	1.30	2.19	5.92
Disallowed (%)	0.00	0.00	0.06

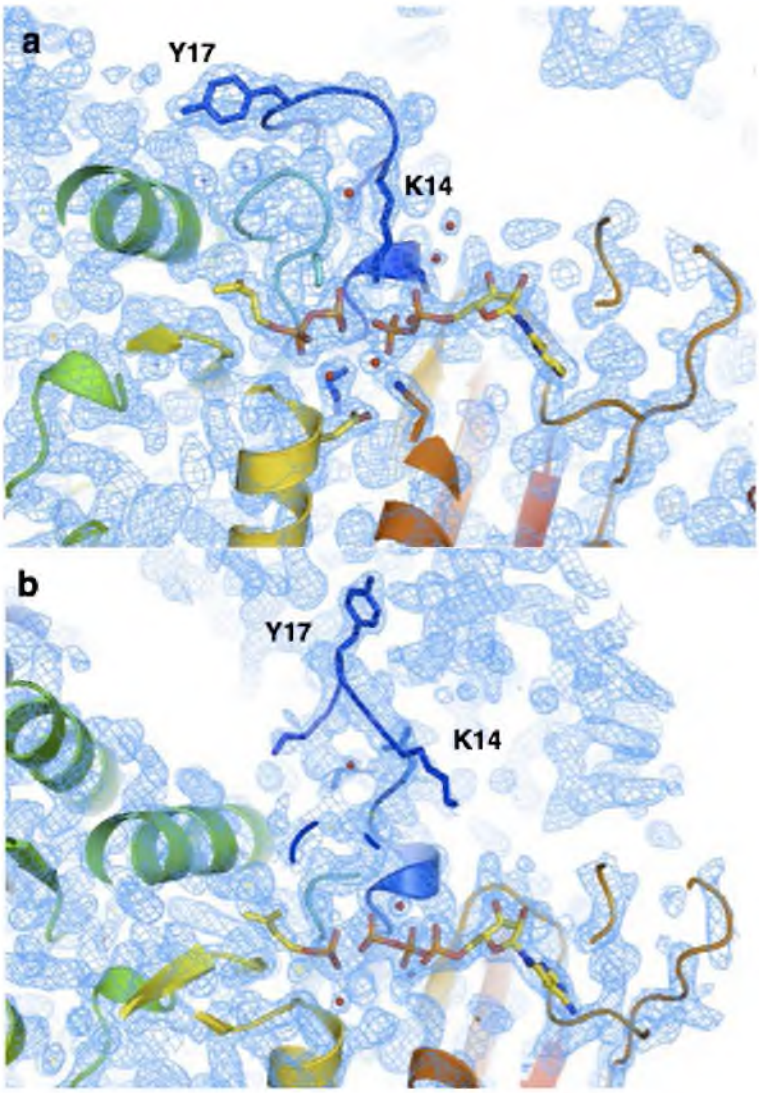


Figure 2.14. $2F_o$-F_c electron density maps of THA IPK showing alternate conformations for the η1-αA loop containing Lys14, shown in blue cartoon. (a) In the first conformation, Tyr17 is oriented towards the long αC helix and Lys14 is in a transition state stabilizing conformation. (b) In the second conformation, the η1-αA loop is turned away from the αC helix and toward the neighboring subunit. The side chain of Lys14 is stabilized in a hydrophic cleft in the neighboring subunit and forces Lys14 away from the phosphorylation reaction site.

movement resulted in a significant shift in the side chain position of Lys14, an invariant residue in IPKs and in FomA with direct involvement in catalysis to be discussed later. Moreover, in molecules C and D, an alternate conformation was found for another conserved catalytic residue, His50 (Figure 2.15). An active site water molecule found in AAK enzymes was found only in molecules A, C and D. The unique characteristics of these asymmetric unit molecules allowed us to choose molecule A as the best structure that represents the THA IPK product complex (IPK•IPP•ADP).

Discussion

Fold and quaternary structure of IPK. IPK belongs to the amino acid kinase (AAK) family of enzymes that also includes NAGK, the AK N-terminal domain, G5K, CK, UMPK and FomA *(37, 38)* (Figure 2.3). This family is characterized by the open αβα sandwich fold first observed in *E. faecalis* CK, and catalyzes the transfer of a phosphoryl group from a nucleotide to a small molecule acceptor. Depending on the functional group that is phosphorylated, this family can be divided into two subdivisions (Figure 2.16). Enzymes in the "carboxylate" subdivision catalyze the transfer of a phosphoryl group to a carboxylate or carbamate group (NAGK, CK, G5K and AK), while those in the "phosphate" subdivision transfer a phosphoryl group to a phosphate or phosphonate (UMPK, FomA, IPK) *(36, 39)*. THA IPK is structurally most similar to FomA, where the backbone RMSD is 2.6 Å on 224 residues. Overlap of THA IPK with FomA (Figure 2.17a) and NAGK (2.17b) also aligns their respective substrates and most of their active site residues, suggesting similar catalytic mechanisms. As expected, MTH IPK shares high structural similarity with THA IPK with RMSD of 1.9 Å over 223 Cα

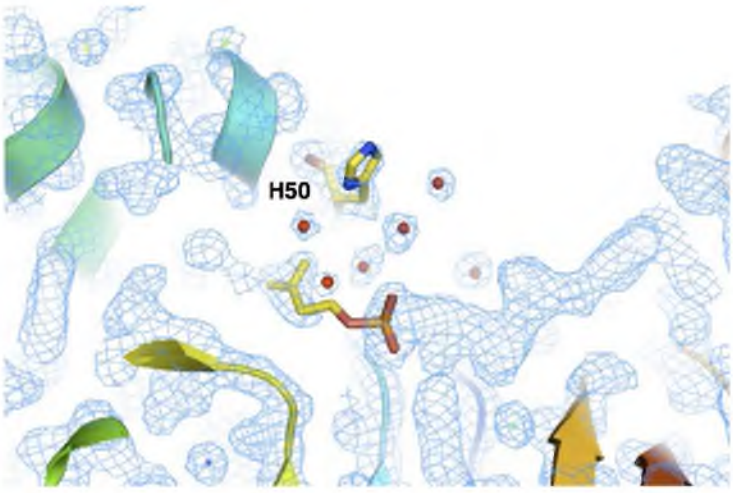


Figure 2.15. A $2F_o$-F_c map showing an alternate conformation found for His50 in one of the subunits in the crystal asymmetric unit containing both the IPK•IPP•ADP and IPK•IP•ATP complexes.

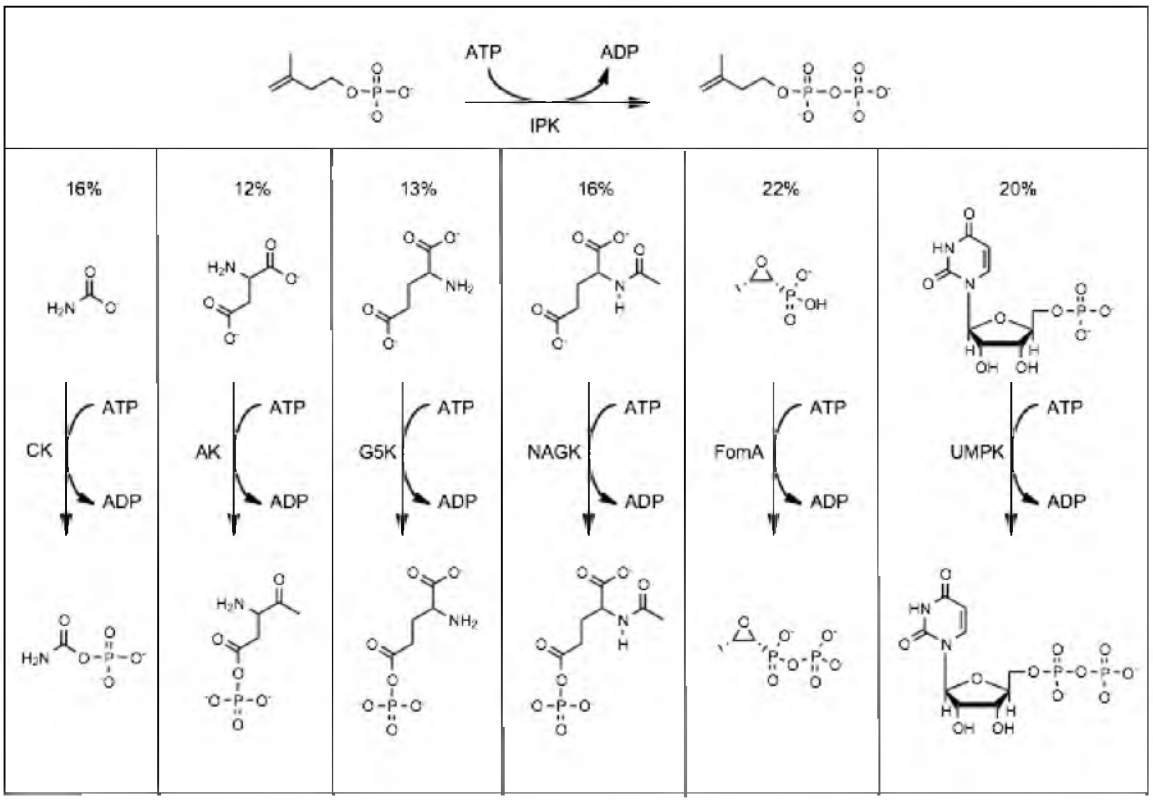


Figure 2.16. Reactions catalyzed by members of the amino acid kinase family of enzymes. Carbamate kinase (CK), aspartokinase (AK), L-glutamate-5-kinase (G5K), and N-acetyl-L-glutamate kinase (NAGK) whose substrates contain either a carboxylate or carbamate group, belong to the carboxylate subdivision. Fosfomycin kinase (FomA), uridine monophosphate kinase (UMPK) and isopentenyl phosphate kinase (IPK) phosphorylate substrates with either a phosphate or phosphonate group and belong to the phosphate subdivision in the AAK family. Enzymes in the family have highly divergent sequences, with IPK and FomA having the highest sequence identity at 20-25%.

Figure 2.17. Stereoviews of the superimposed structures of THA IPK, FomA and NAGK. (a) Superposition of THA IPK and FomA. The top panel shows good overlap of the overall tertiary structures of the two enzymes as well as the location of their substrates. THA IPK is colored green and FomA is colored blue. The substrates of THA IPK (IP and ATP) are colored black while those of FomA (fosfomycin and the inert ATP analog, ANP or AMPPNP) are colored white. Conserved residues that bind the respective substrates of the two enzymes are shown as sticks. (b) Stereoviews of the superimposed structures of THA IPK (green) and NAGK (red). The substrates IP and ATP of THA IPK (black) and those of NAGK (NAG and ANP, grey) superimpose well (top panel), and the residues that bind ATP in both enzymes are conserved. Residues that bind the phosphoryl acceptor (IP and NAG) differ between the two enzymes. The amphipathic molecule IP is bound on the phosphate end with polar residues, while its aliphatic C5 tail is bound by hydrophobic residues. NAG, which contains two carboxylate groups and an acetyl group, is bound in the enzyme through interactions with hydrogen bond donors Asn160, Asn158 and Arg166.

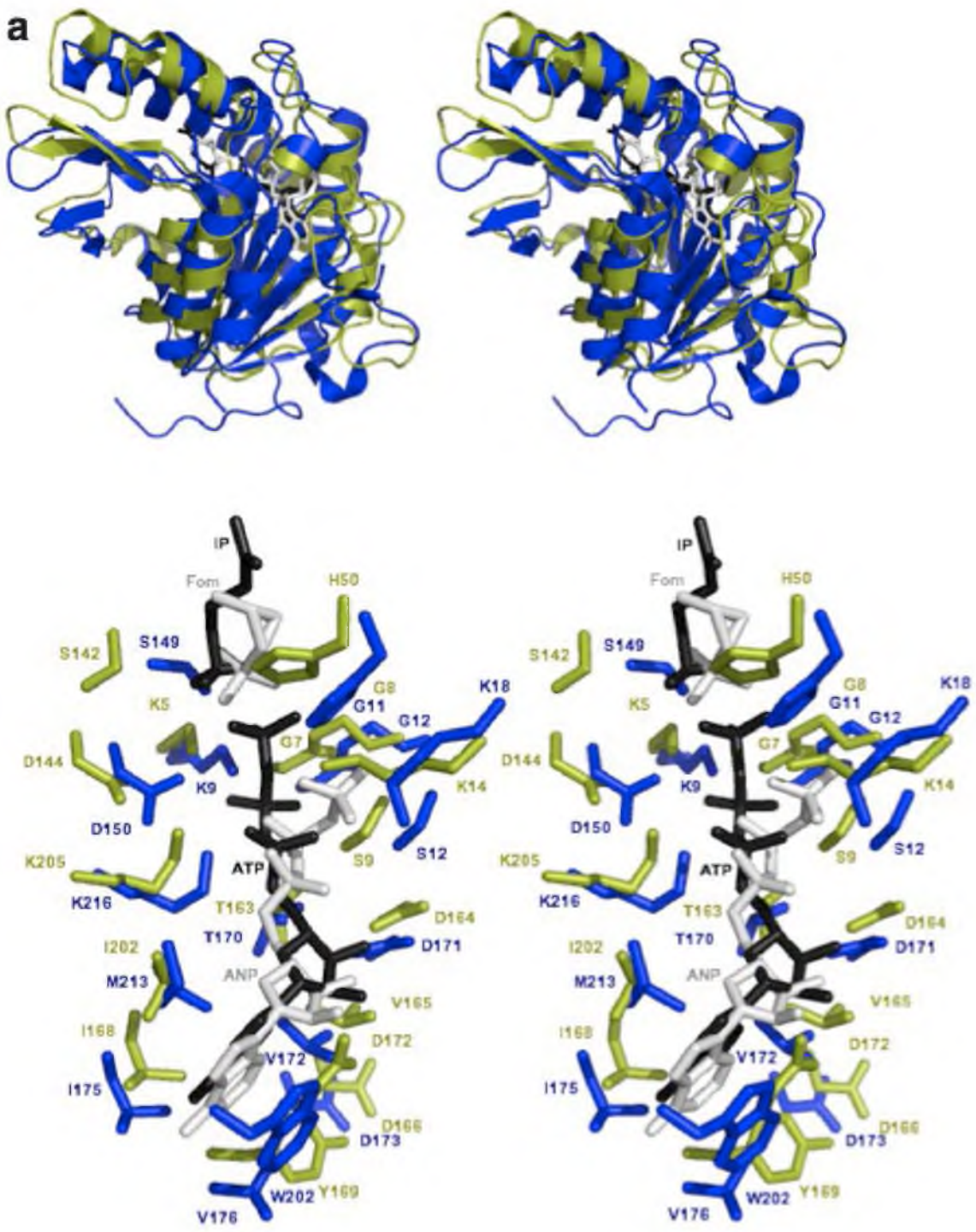


Figure 2.17 continued

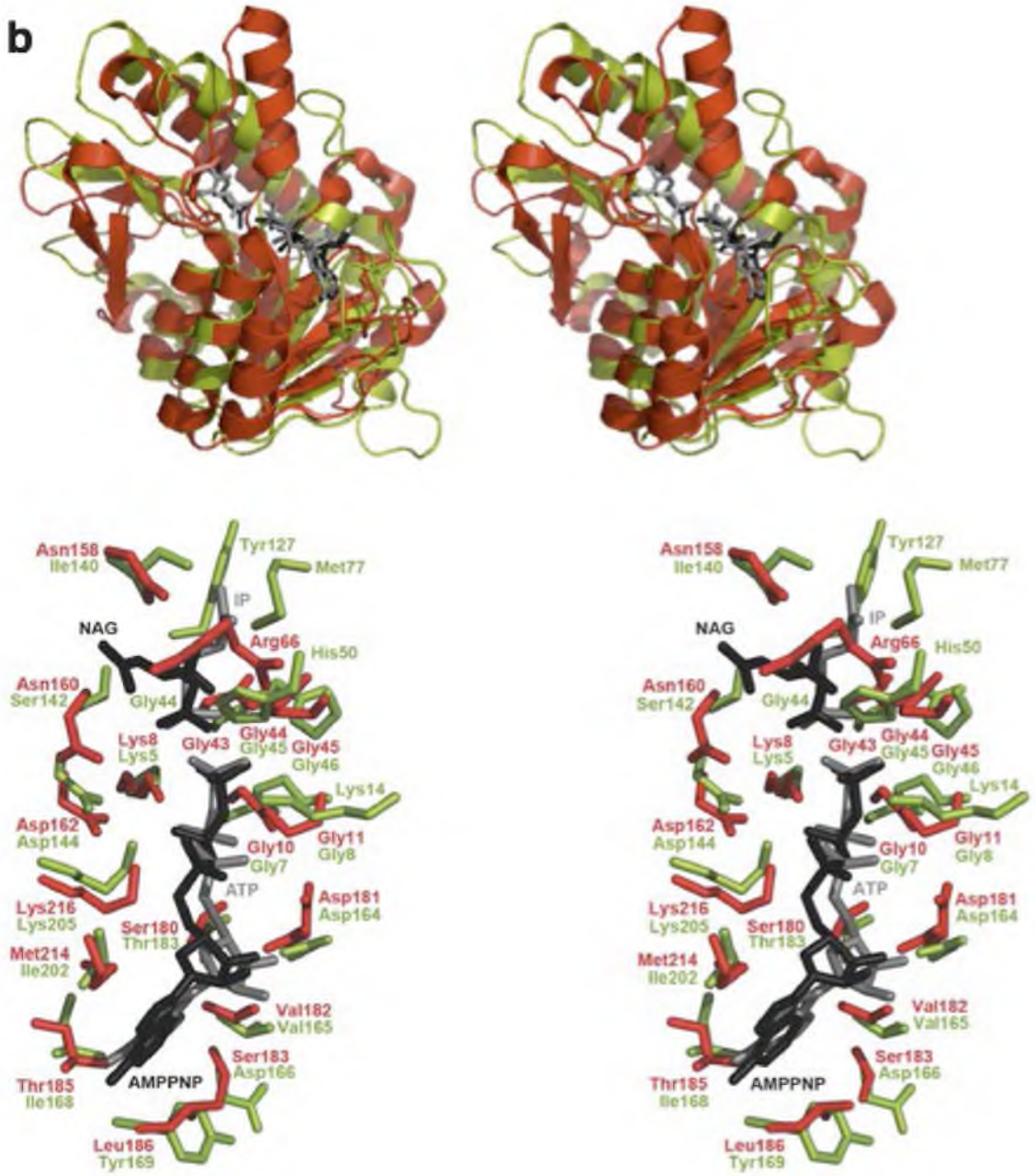


Figure 2.17 continued

atoms, although sequence identity is only 32% (Figure 2.18a and b). These structural comparisons confirm the correct assignment of THA and MTH IPKs to the AAK family of enzymes.

The THA IPK structure contains an N-terminal domain (residues 1–155) that binds IP and a C-terminal domain (residues 156–245) that binds ATP. Each molecule of THA IPK consists of 16 β-strands, eight α-helices and two 3_{10} helices interconnected by loops and turns (Figure 2.19a). The open αβα sandwich fold formed by these secondary structures is highly similar to that of NAGK: the same structural elements form the central β-sheet core (β5, β8, β2, β1, β11, β15, β16, β14) and the two flanking layers of α-helices (αC, αA, αH on one side and αD,αE, αG, αF on the other). Three β-hairpin turns reside in the N-terminal domain, two of which form the surrounding walls of the aliphatic tail of IP (β3-β4, β9-β10) and another forming one side of the IP binding pocket (β6-β7). The β3-β4 hairpin turn of THA IPK is shorter than that of NAGK but is in the same closed conformation that allows it to cap the IP binding site and shield its hydrophobic surroundings from solvent molecules *(19)*. The remaining β12-β13 hairpin turn resides in the C-terminal domain and hovers above the purine ring of ATP. This β hairpin is found within the segment that connects the core β-sheet strands β11 and β14 and is longer than the corresponding segment in NAGK. The longest helix in THA IPK, αC, aligns well with αC in NAGK, but is one turn longer (Figure 2.19b). The segment between β14 and αG (residues 189–201) in the THA IPK structure is disordered in all molecules in the asymmetric unit, as evidenced by uninterpretable electron density. This segment contains the helix, αF and the αF-αG junction loop that form one side of the ATP binding site (Figure 2.19b). The corresponding segment in MTH IPK is missing

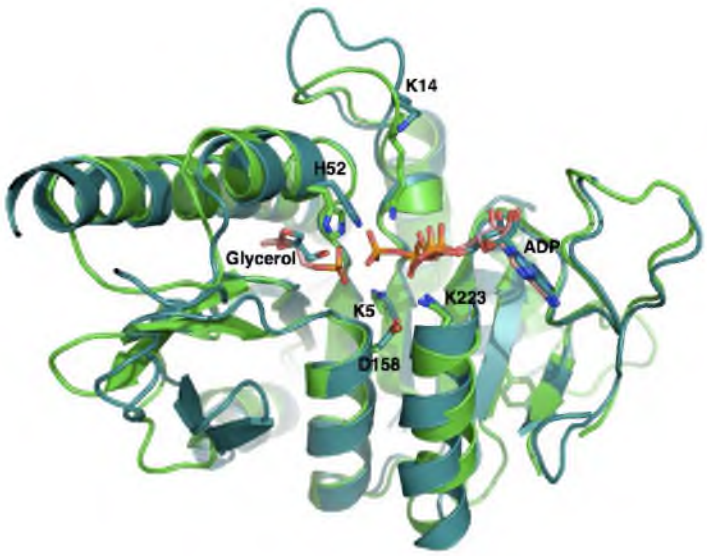


Figure 2.18. Superposition of THA IPK (green) and MTH IPK (deep teal) complexes. Catalytic residues, including the lysine triangle, the conserved histidine and the organizing aspartic acid side chain are shown in sticks. Both structures are missing the αF helix and the loop connecting the αF and G helices, which contain the conserved DVTGG sequence that are known to bind both substrate and nucleotide in FomA. The αC helix of MTH IPK is one turn longer than that of THA IPK.

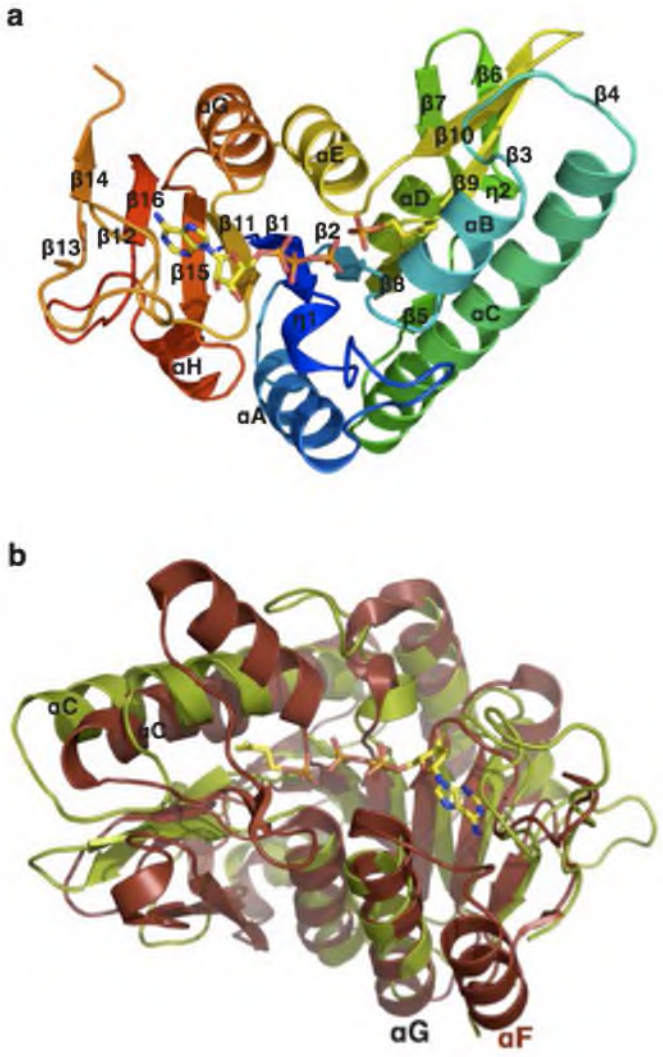


Figure 2.19. Tertiary structure of THA IPK compared with that of NAGK. (a) THA IPK with substrates IP and ATP in the active site, showing secondary structures that bind these ligands. (b). Superposition of THA IPK (green) and NAGK (red) structures, showing the missing αF helix and the loop that connects it to the αG helix. The αC helix of THA IPK is one turn longer than that of NAGK.

(204–218), as is also seen in one molecule of G5K (2J5T) *(22)*, suggesting that the flexibility of these secondary structures might be important for the binding and release of ATP.

The THA and MTH IPKs are dimeric in the crystal, consistent with the observed oligomerization state during gel filtration chromatography (Figure 2.20a). The dyadic axis is perpendicular to the core β-sheet that runs across the dimer (16 β-sheets, 8 per subunit). As in all enzymes of this family that catalyze phosphoryl transfer, each active center is confined to a single subunit. However, dimer formation may be structurally significant since N-terminal domain loops surrounding the active site are anchored to the other subunit of the dimer *(19)*. Analysis of the dimerization surface of THA IPK using the EBI PISA *(40)* server resulted in a complexation significance score of 0.797, corresponding to a tight dimer that buries a total of 3275 $Å^2$ of solvent accessible area between the two monomers. The compactness arises from a total of nine hydrogen bonding interactions between residues in the long interfacial αC helices of both subunits and between the αD helix of one subunit and the β4, β6, and β9 strands and the η2 helix of the other. Eight electrostatic interactions between the αC helices of the two subunits and between the β6 strand of one subunit and the β7 strand of the other further stabilize the dimer. Finally, hydrophobic interactions mediated by residues in the αC, αD and η3 helices, and the β4, β5, β6, β9 and β10 strands further bind the two molecules. Table 2.2 a and b list all electrostatic, hydrophobic and hydrogen bonding interactions that stabilize the dimer.

Although essentially the same secondary structures are used by the AAK enzymes in dimerization, different quaternary architectures are generated by changing the angle

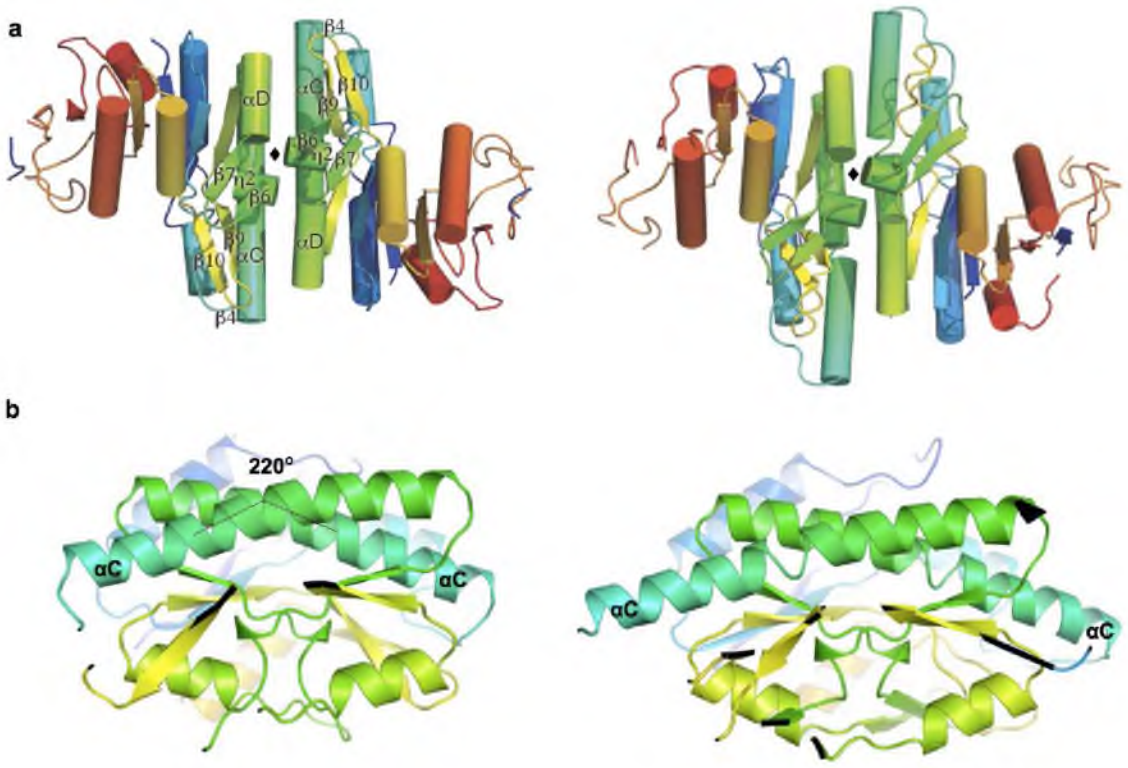


Figure 2.20. THA IPK and MTH IPK dimers. (a) The dimer of THA IPK showing secondary structures containing residues that interact at the dimerization surface. The same secondary structures form the dimeric interface in MTH. The dyadic axis is depicted by (♦). The unique dimeric architectures of THA and MTH IPKs are generated by the crossover angle between the αC helices of each subunit. In both enzymes, this angle is 220°.

Table 2.2 Noncovalent interactions on the dimerization surface of THA IPK (a) and MTH IPK (b). Data on this table were obtained from analyses of the crystal structures using the EBI PISA server.

a. THA IPK

Hydrogen bonds

Chain B	Distance (Å)	Chain A
Y117 (OH)	2.89	Y70 (OH)
S97 (OG)	3.01	E78 (OE1)
R75 (NH2)	2.86	D86 (OD1)
S71 (OG)	2.64	P95 (O)
Y117 ()H)	2.68	S101 (O)
R104 (NH2)	3.82	T112 (OG1)
Y70 (OH)	2.77	Y117 (OH)
R116 (NH2)	2.74	I132 (O)
R94 (NH1)	3.65	Y138 (OH)
E78 (OE1)	3.57	S97 (OG)
D86 (OD1)	2.65	R75 (NH2)
P95 (O)	2.89	S71 (OG)
S101 (O)	2.60	Y117 (OH)
Y105 (O)	2.86	R116 (NH1)
A120 (O)	3.10	K62 (NZ)
I132 (O)	3.14	R116 (NH2)
Y138 (OH)	3.54	R94 (NH1)

Salt bridges

Chain B	Distance (Å)	Chain A
R75 (NH2)	2.86	D86 (OD1)
R75 (NE)	3.88	D86 (OD1)
R104 (NE)	3.33	D110 (OD2)
R104 (NH1)	3.78	D110 (OD2)
D86 (OD1)	2.65	R75 (NH2)
D86 (OD1)	3.76	R75 (NE)
D110 (OD1)	3.90	R104 (NE)
D110 (OD2)	3.82	R104 (NH1)

Table 2.2 continued

b. MTH IPK

Hydrogens bonds

Chain B	Distance (Å)	Chain A
S104 (OG)	3.63	Q81 (OE1)
K85 (NZ)	2.80	N88 (OD1)
N82 (ND2)	3.26	S89 (O)
K85 (NZ)	3.74	C92 (SG)
N82 (ND2)	2.89	D93 (OD1)
K86 (NZ)	2.84	D93 (OD1)
Q106 (NE2)	3.11	S104 (OG)
Y129 (OH)	2.69	S108 (O)
R112 (NH2)	3.20	S124 (O)
R112 (NH2)	3.78	S124 (OG)
R112 (NH2)	3.14	S128 (OG)
R112 (NH2)	2.90	S128 (OG)
R74 (NH1)	3.65	G133 (O)
Q81 (OE1)	3.62	S104 (OG)
N88 (OD1)	2.66	K85 (NZ)
S89 (O)	3.28	N82 (ND2)
C92 (SG)	3.72	K85 (NZ)
D93 (OD1)	2.71	N82 (ND2)
D93 (OD1)	2.88	K86 (NZ)
S104 (OG)	3.46	Q106 (NE2)
S108 (O)	2.68	Y129 (OH)
S124 (OG)	3.34	R112 (NH2)
S128 (OG)	3.28	R112 (NH2)
S128 (OG)	3.06	R112 (NH2)
E132 (O)	3.00	R74 (NE)
E132 (OE2)	3.31	K151 (NZ)
G133 (O)	3.73	R74 (NH1)

Salt bridges

Chain B	Distance (Å)	Chain A
K86 (NZ)	2.84	D93 (OD1)
K86 (NZ)	3.51	D93 (OD1)
K151 (NZ)	3.27	E132 (OE2)
D93 (OD1)	2.88	K86 (NZ)
D93 (OD2)	3.05	K86 (NZ)
E132 (OE1)	3.20	K151 (NZ)
E132 (OE2)	3.31	K151 (NZ)

between the long interfacial helices αC of the dimer subunits *(41)*. In the case of THA IPK, the crossover point of these helices is found on the fifth turn from the N-terminus, and the rotation angle around an axis that penetrates the inter-subunit interface perpendicularly at this point is ~220° (Figure 2.20b). In MTH IPK, the crossover point shifts to the sixth turn from the N-terminus while the rotation angle remains at ~220° (Figure 2.20b). In NAGK and CK the crossover points in αC occur in the third turn, and shifts to the fifth turn in G5K, while the rotation angles are 110° in both NAGK and CK and 260° in G5K *(41)*.

The active site. The THA IPK structures presented here are the first to show the molecular details of binding IP and IPP molecules by an enzyme (Figure 2.21a and b). Many characteristics of AAK enzymes have been described previously in the discussion of the NAGK structure, and several of these traits are conserved in both THA and MTH IPKs. Briefly, the IP (IPP) binding pocket is located in the N–terminal domain formed by the β2–αB loop, the αB helix, the β3–β4 hairpin, the N–terminus of the αC helix, the β8–β9 loop, the β9–β10 hairpin turn, and the β10–αE loop. The αB helix contributes to binding IP (IPP) through the orientation of the positive end of the helix dipole and the hydrogen bonding and electrostatic interactions made by the charged His50 residue. In the FomA structure, the αB helix is ordered in the presence of fosfomycin and disordered in its absence (Figure 2.21c) *(37)*. Comparison of AAK enzyme structures revealed that only the enzymes in the "phosphate" subdivision (UMPK, FomA and IPK) align their respective αB helices. Moreover, an aligning residue at the His50 position is found exclusively in these enzymes, including UMPK *(42)* (Figure 2.21d).

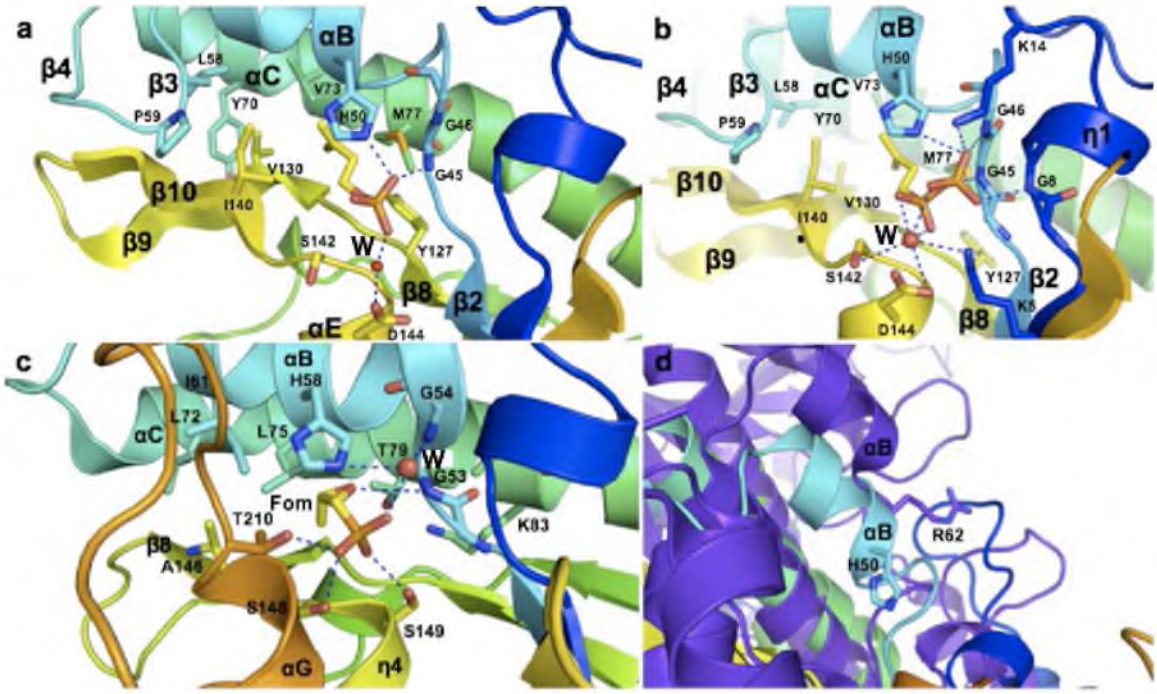


Figure 2.21. Comparison of the substrate binding sites of THA IPK, FomA and UMPK. (a) and (b) The IP and IPP binding sites with important binding and orienting residues shown as sticks. Water molecules (W) are shown as red spheres. The water molecule in the IPK•IPP•ADP complex has six potential hydrogen bonding interactions. Hydrophobic residues bind the C5 tail of both IP and IPP. The phosphate groups of IP and IPP are bound by His50, Ser142, the amino group of Gly45, and a conserved water molecule. This water bridges the phosphate group of IP to Asp144 via hydrogen bonding. His50 binds the β-phosphate of IPP. This movement from the terminal phosphate of IP to the β-phosphate of IPP is one of the few coordinate shifts observed in the IPK•IPP•ADP complex. Tyr127 and Met77, found immediately below the aliphatic tail of IP, enhance the hydrophobicity of the IP binding site and are absent in the active site of FomA (c) His58 of FomA interacts with the phosphonate group of fosfomycin via a bridging water molecule. Ser148, Ser149 and Thr210 orient the phosphonate group of fosfomycin for nucleophilic attack. (d) Superposition of the αB helices of THA IPK and UMPK, showing the alignment of these secondary structures. Arg62 of UMPK corresponds to His50 of THA IPK.

His50 in IPK•IP•ATP poises IP for nucleophilic attack on P_γ of ATP by hydrogen bonding with a nonbridging oxygen atom in the phosphate of IP. In the IPK•IPP•ADP structure, this residue moves slightly to make a hydrogen bond with a nonbridging oxygen atom in the IPP terminal phosphate (Figure 2.21a and b), facilitating the reverse reaction. In fact, in comparing IPK•IP•ATP and IPK•IPP•ADP (RMSD 0.4 Å), the slight change in the position of His50 to hydrogen bond with the terminal phosphate of the product is one of the very few coordinate shifts that were observed. In FomA, His58 positions fosfomycin for nucleophilic attack on the phosphate donor through hydrogen bonding with an intervening water molecule *(37)* (Figure 2.21c*)*. The corresponding residue in *E. coli* UMPK, Arg62, hydrogen bonds with the terminal phosphates of both substrate and product and stabilize P_γ of ATP through electrostatic interactions *(42)*. These observations together with the invariance of this histidine residue in putative IPK homologues *(24, See Supporting Information)*, the apparent pKa of 6.5 for THA IPK (6.8 for MTH IPK) from pH dependence studies *(24)*, suggest that this His50 may perform direct catalytic roles similar to those proposed for Arg62 in UMPK, besides being a signature trait of IPKs *(42)*. Independent experiments performed by Dellas and Noel showed that the corresponding residue in MJ IPK, His60 plays an important role in both substrate and product sequestration *(39)*. Isosteric mutations on His60 of MJ IPK in which the $N\varepsilon_1$ and $N\varepsilon_2$ atoms were mimicked through His60Asn and His60Gln mutations, respectively, showed that the protonated $N\varepsilon_2$ nitrogen atom accomplishes the said role by donating a hydrogen bond to a nonbridging O atom of the terminal phosphate group of either IP or IPP. The His60Gln mutant had a K_M^{IP} of 34.5 μM compared to 4.30 μM for the wild-type MJ IPK, suggesting that His60 is important for ground state

binding. The additional flexibility of a Gln residue relative to His may hinder its ability to bind IP as effectively as wild-type IPK. Moreover, the k_{cat}/K_M value for the His60Gln mutant was more than 300 times lower than that of the wild-type IPK, suggesting that the added charge and lowered conformational stability of His60 are important for the stabilization of the highly negatively charged transition state during phosphoryl transfer. His60Ala and His60Asn mutants of MJ IPK were both inactive *(39)*.

The specificity of IPK arises from hydrophobic residues that cradle the aliphatic portion of IP and a constellation of conserved residues that hydrogen bond with its phosphate moiety (Figure 2.21a). Residues Tyr70, Met77 and Val73 (αC N-terminus), Val130 and Ile140 (β9-β10 hairpin) and Ala53 (αB C-terminus) and Tyr127 (β8-β9 loop) all form the hydrophobic pocket for the apolar tail of IP. In addition, Leu58 and Pro59 of the β3-β4 hairpin enhance the hydrophobic surroundings by capping the IP binding pocket. Notably, the β3-β4 hairpin is found only in the structures of NAGK and IPK. This motif is in the closed conformation in substrate-bound NAGK structures *(X)* as in the IPK•IP•ATP and IPK•IPP•ADP structures. The side chains of some, if not all, of these hydrophobic residues may give rise to the chain length specificity of IPK. Geranyl phosphate, which is one isoprene unit longer than IP, is a poor substrate for IPK, indicating that the IP binding site cannot easily accommodate a C_{10} isoprenoid chain *(24)*. Thus, the mutation of the residues in the IP binding site may expand this space and accommodate different conformations of longer chain isoprenoids or other bulkier substrates.

The phosphate moiety of IP is recognized in the active site by structural motifs that are well conserved in the AAK family. In IPK•IP•ATP, the three nonbridging

oxygen atoms of IP make hydrogen bonding interactions with the $N\varepsilon_2$ of His50, the N atom of Gly45 in the β2-αB loop, and Asp144 through a bound water molecule (Figure 2.21a). This water molecule is found in the holoenzyme structures of E. coli NAGK (pdb ID 1GS5), P. furiosis UMPK (pdb ID 2BMU), E. coli UMPK (pdb ID 2BNE) and FomA (pdb ID 3D41). A water molecule in a similar position is present in IPK•IPP•ADP (Figure 2.21b). The β2-αB loop, which contains Gly45, is one of the conserved glycine-rich loops in the AAK family and is thought to stabilize the transition state intermediate and the phosphorylated product of these enzymes *(37, 43)*. In IPK•IPP•ADP, Gly45 of this loop makes a hydrogen bond with the P_α-P_β bridging oxygen atom of IPP. The other glycine-rich loop, β1-η1, contains Gly8 that stabilizes IPP by making a hydrogen bond with a nonbridging oxygen atom of its terminal phosphate (Figure 2.21b).

The β10–αE loop completes the structural elements involved in binding the polar end of IP. This loop contains Ser142, which is conserved in the "phosphate" subdivision (Thr120 in *P. furiosis* UMPK, Ser149 in FomA) *(23, 37)* and makes a hydrogen bond with the terminal phosphate of the substrates. In IPK•IP•ATP and IPK•IPP•ADP, the orientation of the side chain hydroxyl group of Ser142 does not allow a hydrogen bond with substrates or products (Figure 2.21a and b). However, an alternate rotamer could provide hydrogen bond stabilization as the substrates moved to the transition state. The positive end of the helix dipole of αE, initiated by the β10-αE loop, may further stabilize the polar head of the IP molecule as well as the negatively-charged transition state intermediate, as in other structures of AAK enzymes.

Thus, the specificity of IPK towards IP arises from a combination of hydrophobic interactions with the hydrocarbon tail and hydrogen bonding and electrostatic interactions

with the phosphate headgroup to match the amphiphatic character of IP. This differentiates the IPK active site from other enzymes in the family whose substrates do not have this intrinsic amphiphathicity, such as that of NAGK whose active site is laden with both polar and charged residues that interact with the acetyl and carboxyl groups of NAG *(36)* (Figure 2.17b).

The location of the ATP binding site in both THA and MTH IPKs superimposes well with those of NAGK and other AAK enzymes. Conserved amino acid residues in these enzymes, which have previously been shown to form hydrogen bonding, electrostatic and hydrophobic interactions with the purine ring, sugar and polyphosphate moieties of ATP, are also found in IPK. For instance, the NAGK residues Met214 (purine stacking), Asp181 (ribose 3'-OH binding), Asp162 (polyphosphate positioning) and Lys8 (phosphate hydrogen bonding) are also conserved in THA IPK as Ile202, Asp164, Asp144 and Lys5, respectively *(19, 36)* (Figure 2.17b). In MTH IPK, the corresponding residues are Met220, Asp178, Asp158 and Lys5 (Figure 2.22). Other ATP-binding residues in THA IPK that superimpose with similar ATP-binding residues in NAGK are shown in (Figure 2.17b). The β1-η1 glycine-rich loop that contains Gly8 engages the P_β and P_γ of ATP in IPK•IP•ATP through hydrogen bonding (Figure 2.21a). These interactions by the conserved Gly8 with the phosphate groups ATP (and with the terminal phosphate of IPP, as described earlier) are seen for the corresponding residue in other enzymes and reflect a role in phosphoryl transfer *(42, 43)*. Residues Asp172 and Lys174 form hydrogen bonds with the 2'- and 3'-OH groups of ATP, respectively. Asp172 and its corresponding residues in other AAK enzymes may control selectivity for ATP over 2'-deoxy-ATP (dATP). Finally, a hydrogen bond with the side chain carboxylate of

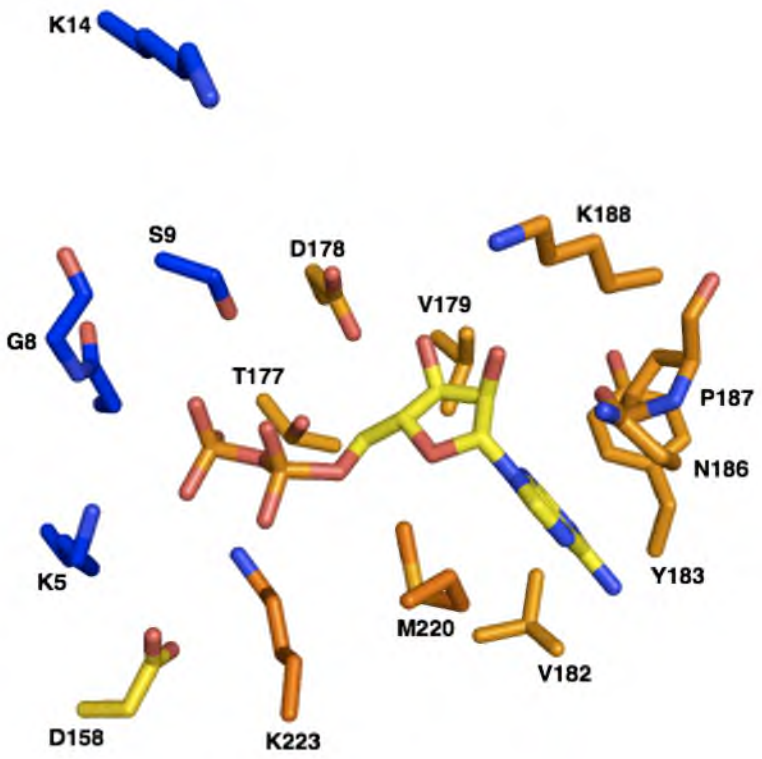


Figure 2.22. The nucleotide binding site of MTH IPK contains residues that are conserved in THA IPK, NAGK and other amino acid kinase family enzymes.

Asp164 completes an intricate network of interactions that secure this end of ATP, thus positioning the polyphosphate end for nucleophilic attack by IP (Figure 2.23).

As stated earlier, αF and the loop that connects it to αG were not modeled in our IPK structures. These secondary structures contain Phe191, analogous to Trp 202 in FomA, which stacks on the adenine ring, and the conserved DVTGG sequence that makes important binding interactions with the nucleotide, as seen in other AAK enzymes. We suspect that these interactions are largely conserved in IPK, although a more complete IPK structure would be necessary to unequivocally establish the nature of these interactions. The flexibility and location of these absent secondary structures suggest a role in nucleotide sequestration and release.

The IPK•IP•ATP structure was obtained without resorting to a substitution of the P_β-P_γ bridging oxygen atom that would prevent phosphoryl transfer. In addition, phosphoryl transfer between added substrates during the crystallization experiments provided IPK•IPP•ADP. These two structures, providing snapshots of the IPK enzyme before and after catalysis, suggest an important role for Lys14 in stabilizing the transition state (Figure 2.10b and Figure 2.11c). This residue is also conserved in FomA and homologs of IPK *(24, See Supporting Information)*. This transition state-stabilizing role of Lys14 in THA IPK was confirmed by kinetic studies on a Lys14Ala mutant of THA IPK. This mutant has a K_M^{IP} of 3.4 μM similar to that for wild-type IPK, suggesting that Lys14 does not directly interact with the substrate IP. On the other hand, K_M^{ATP} for the Lys14Ala mutant is almost twice that of wild-type THA IPK, while its k_{cat} is 460-fold lower than wild type (Figure 2.24). These support the putative roles of Lys14 inferred from the crystal structure of THA IPK, in which it makes hydrogen bonding interactions

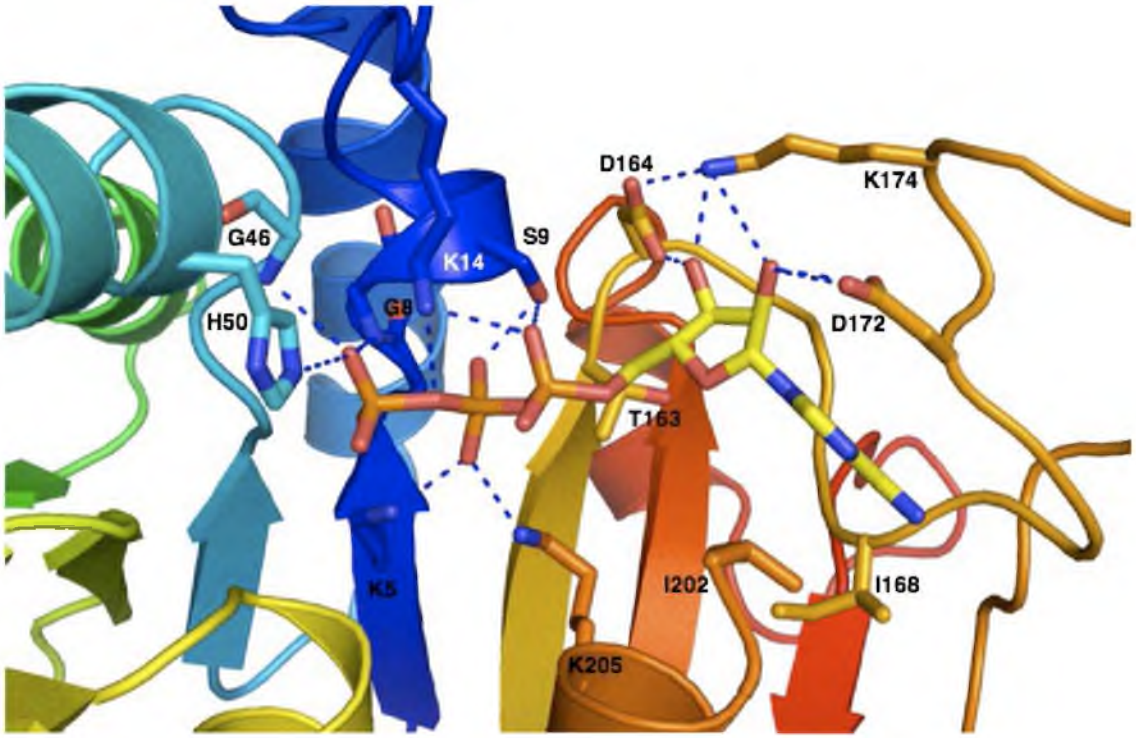
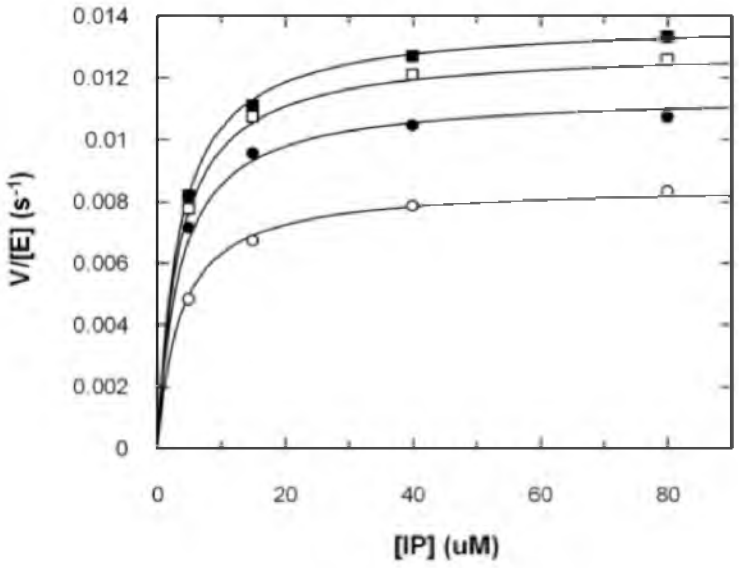


Figure 2.23. The ATP binding site of THA IPK showing residues that participate in hydrogen bonding and hydrophobic interactions. Hydrogen bonds are shown as blue dashed lines. The Mg^{2+} ion is not found in THA IPK structures, but are presumed to coordinate three nonbridging oxygen atoms of the ATP phosphate moiety.

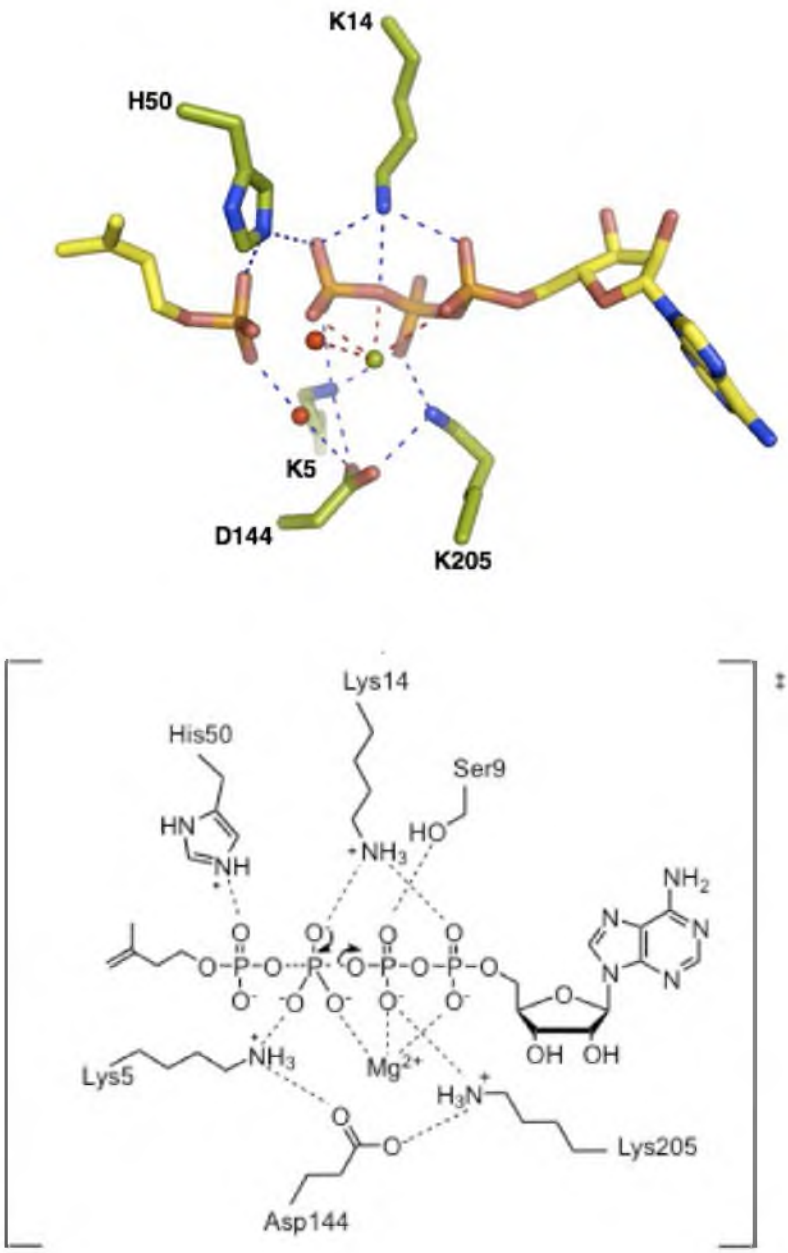


	Wild-type IPK	Lys14Ala IPK
k_{cat} (s^{-1})	8.0 (± 0.2)	1.8 (± 0.04) × 10^{-2}
K_M^{IP} (μM)	4.4 (± 0.5)	3.2 (± 0.5)
k_{cat}/K_M^{IP} ($M^{-1}s^{-1}$)	1.8 × 10^6	5.5 × 10^3
K_M^{ATP} (μM)	6.0 (± 0.5)	10.5 (± 0.8)
K_d (μM)	4.6 (± 1.5)	3.8 (± 0.2)

Figure 2.24. Michaelis Menten kinetics of the Lys14Ala mutant of THA IPK. This mutation showed that Lys14 is important for transition state stabilization.

to bind the phosphate moiety of ATP as well as help stabilize the highly negatively charged transition state through its positive charge.

An associative mechanism of phosphoryl transfer was proposed for NAGK as well as for its closest homolog, FomA, Structural evidences for an associated mechanism were obtained from structures of the NAGK enzyme containing the transition state mimic AlF_4^-, and in the case of FomA, in complex with fosfomycin vanadate and MgADP *(38, 43)*. The similarity of the overall fold and the superposition of the substrate and nucleotide binding sites of THA and MTH IPKs with these enzymes suggest a similar mechanism for IPK (Figure 2.25). In IPK•IP•ATP, a nucleophilic oxygen atom of the phosphate group in IP is 2.9 Å from the electrophilic P_γ phosphate atom in ATP, and is poised for attack. The angle formed by the nucleophilic oxygen atom of fosfomycin, the electrophilic phosphorous of the γ-phosphate of ATP and the bridging oxygen atom of the β- and γ-phosphates of ATP is 157.8°. In the modeled FomA•fosfomycin•MgATP complex, this angle is 157.9° *(38)*. The products in IPK•IPP•ADP are likewise poised for the reverse reaction, where the nucleophilic oxygen atom in P_β of ADP is 2.3 Å from the electrophilic P_β phosphate atom of IPP, with an angle of attack of 167.9°. As in NAGK, the conserved glycine rich loops β1-η1 and β2-αB of THA IPK, the positive helix dipoles of αB and αE, and the positive charges of Lys5 and Lys205 (which are positioned by Asp144 towards the polyphosphates by hydrogen bonding), create an environment for stabilizing the negatively charged transition state. It is understood that the concentration of these positive charges at the site of transition indicates an associative mechanism for the IPK-catalyzed reaction, which is expected to have a pentacoordinate transition state with a charge of -3. In contrast, a dissociative mechanism will have a


K14
H50
K5
D144
K205
Lys14
His50
Ser9
HN
NH
HO
NH3
NH2
N
N
N
N
O
O
O

metaphosphate transition state with a charge of -1, which will be overcompensated for by the surrounding residues.

A divalent metal ion, Mg^{2+}, further contributes to stabilization by the same mechanism in addition to its orientational effect on the ATP polyphosphate. In IPK•IP•ATP and IPK•IPP•ADP, a Mg^{2+} ion could not be assigned to the electron density that superimposes onto the Mg^{2+} atom in the structure of NAGK. Instead, a water molecule is assigned to this density since the distance of the centroid of the sphere to the relevant polyphosphate oxygen atoms is more consistent with hydrogen bonding than with metal coordination. Nevertheless, product turnover assays using radioactive γ-[^{32}P] ATP and Mg^{2+} confirmed the divalent metal ion dependence of IPK, a well-established characteristic of kinases *(44)* (Figure 2.26a). Thus, Mg^{2+} is very likely located in the same position as the water molecule seen in the crystal structure, based on superposition with the NAGK complex and the recently determined ternary complex of FomA with fosfomycin and MgATP. The coordination sphere of this Mg^{2+} maybe completed by the three non-bridging O atoms of ATP and three nearby water molecules, as seen in the FomA•fosfomycin•MgATP complex *(38)*. In the structures of THA IPK solved, at least one molecule of water is missing from a complete coordination sphere, while the triphosphate moiety of ATP is in a slightly different conformation that allows one bridging O atom to form a coordinate bond with the divalent metal. It is also possible that the entire enzyme contracts to bring the carboxylate group of Asp150 closer to the divalent metal, thus completing the Mg^{2+} coordination sphere. More THA IPK structures are necessary to confirm these hypotheses. Of the seven structures of FomA deposited in the PDB, only four contained Mg^{2+} in the active site, namely the complexes with

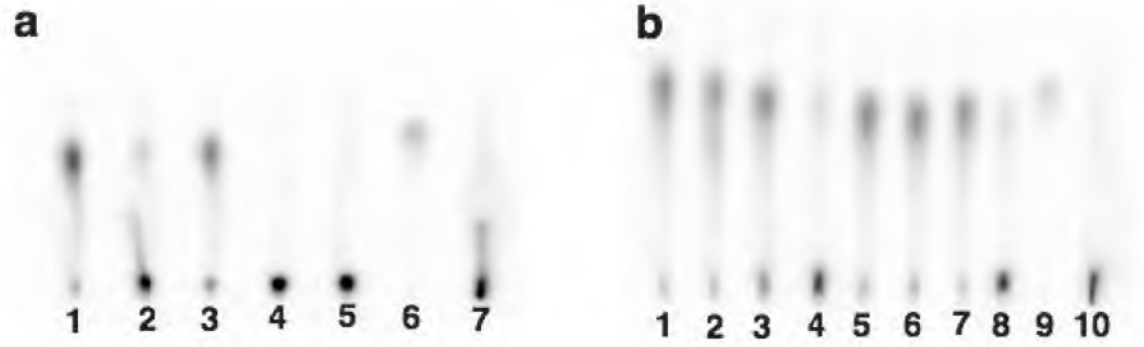


Figure 2.26. Autoradiography assay showing divalent metal ion dependence of the IPK-catalyzed phosphorylation reaction. (a) Mg^{2+} is required for IPK activity. (1) Mg^{2+} present, (2) Mg^{2+} not present, (3) Mg^{2+} present with ten-fold concentration of EDTA, (4) Mg^{2+} not present with ten-fold concentration of EDTA, (5) Mg^{2+} present, IPK not present, (6) [^{14}C] IPP control, (7) γ-[^{32}P] ATP control. (b) IPK is active in the presence of other divalent metals. (1) Mg^{2+}, (2) Mn^{2+}, (3) Zn^{2+}, (4) Cu^{2+}, (5) Ni^{2+}, (6) Cd^{2+}, (7) Co^{2+}, (8) no divalent metal present, (9) [^{14}C] IPP control, (10) γ-[^{32}P] ATP control.

MgATP, fosfomycin vanadate and MgADP, fosfomycin monophosphate and MgADP, and fosfomycin and MgAMPPNP *(37, 38)*. The complexes with the vanadate derivative and the inert ATP analogue are the most ordered among the crystal structures, indicating a correlation between the presence of Mg^{2+} in the metal site and the ordering of flexible loops. Finally, IPK was also active in the presence of Mn^{2+}, Zn^{2+}, Ni^{2+}, Cd^{2+}, and Co^{2+} (Figure 2.26b).

In IPK•IP•ATP, Lys14 is located within hydrogen bonding distance of the non-bridging oxygen atoms of P_α (3.0 Å) and P_γ (2.8 Å) and the P_β-P_γ bridging oxygen atom (3.1Å) in ATP. In IPK•IPP•ADP, Lys14 makes a hydrogen bond with a nonbridging oxygen atom of P_β (2.7 Å) in IPP and the nonbridging oxygen atoms of P_α (2.7 Å) and P_β (3.0 Å) of ADP. A corresponding residue, Lys18, is found in the FomA•Fom•MgAMPPNP complex *(37)*. In this structure, Lys18 only forms a hydrogen bond with a P_γ oxygen atom in AMPPNP, which in turn, is not properly oriented for nucleophilic attack by fosfomycin. The relative positions of IP and ATP in THA IPK are consistent with phosphoryl transfer. Thus, in THA IPK Lys5, Lys14 and Lys205 form a lysine triangle surrounding the phosphate residues in the substrates of IPKs (Figure 2.27). This lysine triangle was also observed in the FomA•Fom•ATP and FomA•FomVO_3•MgADP complexes. Interestingly, the FomA•FomPO_4•MgADP complex, which is analogous to the IPK•IPP•ADP complex, does not possess an ordered lysine triangle and has been interpreted as an enzyme product complex in the act of preparing for product release *(38)*. Thus, the IPK•IPP•ADP complex may approximate a step in the IPK-catalyzed reaction before the one represented by the FomA•FomPO_4•MgADP complex, wherein the products IPP and ADP are bound in the

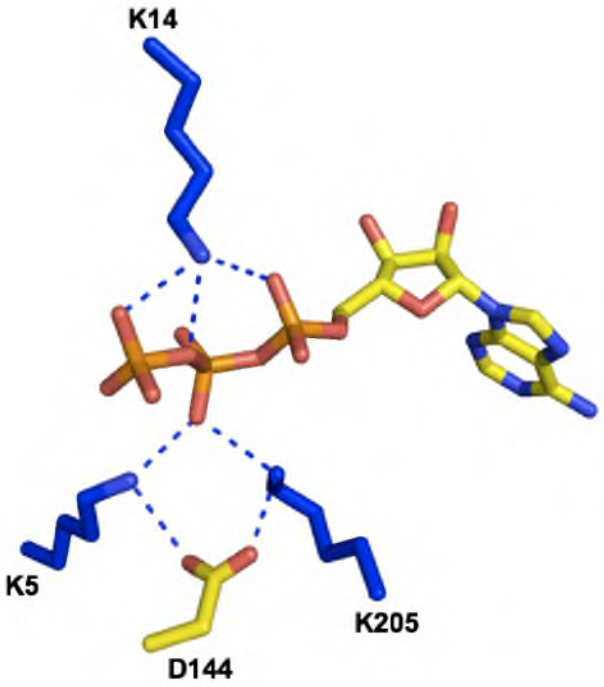


Figure 2.27. The lysine triangle is conserved in IPKs and in FomA. In THA IPK, Lys14 hydrogen bonds with the phosphate moiety of ATP, and forms a triad with Lys5 and Lys205. The residue Asp144 organizes the latter two lysine residues through hydrogen binding interactions. Lys5 and Lys205 each can donate a hydrogen bond to the β-phosphate of ATP. With the position of Lys14, it is presumed to stabilize the transition state during phosphoryl transfer together with Lys5 and Lys205 through neutralization of the negatively charged transition state.

active site just before the movement of the η1-αA loop that contains Lys14, in order to release IPP.

Along with His50, Lys14 is a conserved feature of the IPKs. These residues may have evolved to preferentially stabilize transition states that contain more negative charge than those formed during phosphoryl transfers involving carboxylated and carbamated substrates, thus giving rise to IP kinases. The conservation of these traits in FomA along with other key catalytic residues seen in IPK is also consistent with the acquisition of antibiotic resistance toward fosfomycin in *Streptomyces* by a lateral gene transfer of an ancestral IPK gene from *Archaea*. Chapter 3 presents biochemical and computational studies that elucidate this possible evolutionary relationship between archaeal IPK and bacterial FomA.

References

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome, *Science 298*, 1912-1934.

2. Kornev, A. P., and Taylor, S. S. (2010) Defining the conserved internal architecture of a protein kinase, *Biochim. et Biophys. Acta 1804*, 440-444.

3. Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S. and Sowadski, J. M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase, *Science 253*, 407–414.

4. Herberg, F. W., Doyle, M. L., Cox, S., and Taylor, S.S. (1999) Dissection of the nucleotide and metal-phosphate binding sites in cAMP-dependent protein kinase, *Biochemistry 38*, 6352–6360.

5. Adams, J. A. (2003) Activation loop phosphorylation and catalysis in protein kinases: is there functional evidence for the autoinhibitor model?, *Biochemistry 42*, 601–607.

6. Johnson, D. A., Akamine, P., Radzio-Andzelm, E., Madhusudan, M., and Taylor,

S.S. (2001) Dynamics of cAMP-dependent protein kinase, *Chem. Rev. 101*, 2243–2270.

7. Johnson, L. N., and Lewis, R. J. (2001) Structural basis for control by phosphorylation, *Chem. Rev. 101*, 2209–2242.

8. Nolen, B., Taylor, S. S., and Ghosh, G. (2004) Regulation of protein kinases; controlling activity through activation segment conformation, *Mol. Cell 15*, 661–675.

9. Kornev, A. P., Taylor, S. S., and Ten Eyck, L. F. (2008) A helix scaffold for the assembly of active protein kinases, *Proc. Natl. Acad. Sci. U. S. A. 105*, 14377–14382.

10. Emrick, M. A., Lee, T., Starkey, P. J., Mumby, M. C., Resing, K. A., and Ahn, N. G. (2006) The gatekeeper residue controls autoactivation of ERK2 via a pathway of intramolecular connectivity, *Proc. Natl. Acad. Sci. U. S. A. 103*, 18101–18106.

11. Elphick, L. M., Lee, S. E., Child, E. S., Prasad, A., Pignocchi, C., Thibaudeau, S., Anderson, A. A., Bonnac, L., Gouverneur, V., and Mann, D. J. (2009) A quantitative comparison of wild-type and gatekeeper mutant cdk2 for chemical genetic studies with ATP analogues, *Chembiochem 10*, 1519–1526.

12. Johnson, D. A., Akamine, P., Radzio-Andzelm, E., Madhusudan, M., and Taylor, S. S.(2001) Dynamics of cAMP-dependent protein kinase, *Chem. Rev. 101*, 2243–2270.

13. Liu, Y., Shah, K., Yang, F., Witucki, L., and Shokat, K. M. (1998) A molecular gate which controls unnatural ATP analogue recognition by the tyrosine kinase v-Src, *Bioorg. Med. Chem. 6*, 1219–1226.

14. Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J., and Daley, G. Q.6. Johnson, D. A., Akamine, P., E. Radzio-Andzelm, Madhusudan and S.S. Taylor. (2001) Dynamics of cAMP-dependent protein kinase, *Chem. Rev.101*, 2243–2270.

15. L.N. Johnson and R.J. Lewis (2008) Structural basis for control by phosphorylation, Chem. Rev. 101, 2209–2242.Activation of tyrosine kinases by mutation of the gatekeeper threonine, *Nat. Struct. Mol. Biol. 15*, 1109–1118.

16. Buss, K. A., Cooper, D. R., Ingram-Smith, C., Ferry, J. G., Sanders, D. A., Hasson, M.S. (2001) Urkinase: structure of acetate kinase, a member of the ASKHA superfamily of phosphotransferases, *Journal of Bacteriology 183*, 680-686.

17. Banks, R. D., Blake, C. C. F., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., Merrett, M., and Phillips, A. W. (1979) Sequence, structure and activity of

phosphoglycerate kinase: a possible hinge-bending enzyme, *Nature 279*, 772-777.

18. Waldrop, G. L., Rayment, I., and Holden, H. M. (1994) Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase, *Biochemistry 33*, 10249-10256.

19. Ramon-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I., and Rubio, V. (2002) Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase family, during catalysis, *Structure 10*, 329–342.

20. Marina, A., Alzari, P.M., Bravo, J., Uriarte, M., Barcelona, B., Fita, I., and Rubio, V. (1999) Carbamate kinase: new structural machinery for making carbamoyl phosphate, the common precursors of pyrimidines and arginines, *Protein Science 8*, 934–940.

21. Liu, X., Pavlovsky, A.G., and Viola, R. E. (2008) The structural basis for allosteric inhibition of a threonine-sensitive aspartokinase. J. Biol. Chem. 283, 16216–16225.

22. Marco-Marin, C., Gil-Ortiz, F., Perez-Arellano, I., Cervera, J., Fita, I., and Rubio, V. (2007) A novel two-domain architecture within the amino-acid kinase enzyme family revealed by the crystal structure of Escherichia coli glutamate 5-kinase, *J. Mol. Biol. 367*, 1431–1446.

23. Marco-Marin, C., Gil-Ortiz, F., and Rubio, V. (2005) The crystal structure of Pyrococcus furiosis UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis, *J. Mol. Biol. 352*, 438–454.

24. Chen, M., and Poulter, C.D. (2010) Characterization of thermophilic and archaeal isopentenyl phosphate kinases, *Biochemistry 49*, 207–217.

25. Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures, *Prot. Exp. and Purif. 41*, 207-234.

26. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and Bairoch A. (2005) Protein Identification and Analysis Tools on the ExPASy Server, in The Proteomics Protocols Handbook (Walker, J., Ed.), pp. 571–607.

27. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, Methods in Enzymology, Macromolecular Crystallography (Carter Jr. , C.W. and Sweet, R.M., Eds.), Vol. 276, part A, pp 307–326, Academic Press, New York.

28. Adams, P.D., Afonine P.V., Bunkóczi, G., Chen V.B., Davis, I.W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J.,

Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Cryst. D66, 213–221.

29. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, Acta Crystallogr. 60, 2126–2132.

30. Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D53*, 240–255.

31. McCoy, A.J., Grosse-Kunstleve, Adams, P.D., Winn, M.D., Storoni. L.C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Cryst. 40*, 658–674.

32. Collaborative Computational Project, Number 4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.

33. Holm, L., and Park, J. (2000) DaliLite workbench for protein structure comparison, *Bioinformatics 16(6)*, 566–567.

34. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J., Kapral, G.J., Wang, X., Murray, L., Arendall III, B., Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007) Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35: Web Server issue, W375–W383.

35. DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA.

36. Mabanglo, M.F., Schubert, H.L., Chen, M., Hill, C.P., and Poulter, C.D. (2010) X-ray structures of isopentenyl phsophate kinase, *ACS Chem. Biol.* 2010, *5(5)*, 517-527.

37. Pakhomova, S., Bartlett, S., Augustus, A., Kuzuyama, T., and Newcomer, M. (2008) Crystal structure of fosfomycin resistance kinase FomA from Streptomyces wedmorensis, *J. Biol. Chem. 283*, 28518–28526.

38. Pakhomova, S., Bartlett, S., Doerner, P.A., and Newcomer, M. (2011) Structural and biochemical insights into the mechanism of fosfomycin phosphorylation by fosfomycin resistance kinase FomA, *Biochemistry, ASAP article.*

39. Dellas, N.P., and Noel, J.P. (2010) Mutation of archaeal isopentenyl phosphate kinase highlights mechanism and guides phosphorylation of additional isoprenoid monophosphates. *ACS Chem. Biol., 5(6)*, 589-601.

40. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774-797.

41. Marco-Marin, C., Gil-Ortiz, F., Perez-Arellano, I., Cervera, J., Fita, I., and Rubio, V. (2007) A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of Escherichia coli glutamate 5-kinase. *J. Mol. Biol. 367*, 1431-1446.

42. Briozzo, P., Evrin, C., Meyer, P., Assairi, L., Joly, N., Barzu, O., and Gilles, A.M. (2005) Structure of Escherichia coli UMP kinase differs from that of other nucleoside monophosphate kinases and sheds new light on enzyme regulation, *J. Biol. Chem. 280*, 25533–25540.

43. Gil-Ortiz, F., Ramon-Maiques, S. Fita, I., and Rubio, V. (2003) The course of phosphorous in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with AlF_4^- transition state mimic, *J. Mol. Biol. 331*, 231–244.

44. Tian, G., Kane, L.S., Holmes, W.D., and Davis. S.T. (2002) Modulation of cyclin-dependent kinase 4 by binding of magnesium (II) and manganese (II), *Biophys Chem 95*, 79–90.

CHAPTER 3

THE *STREPTOMYCES*-PRODUCED ANTIBIOTIC FOSFOMYCIN IS A PROMISCUOUS SUBSTRATE FOR ISOPENTENYL PHOSPHATE KINASE

Introduction

Modern enzymes are commonly understood to be highly specific towards the substrates and chemical transformations that they catalyze. This selectivity underlies catalysis in biology. Moreover, in a living cell where the chemical structures of metabolites are highly redundant and where cross wiring of metabolic pathways exists, enzyme selectivity prevents the conversion of unintended substrates that wastes cellular energy or has other deleterious effects on the organism *(1)*. In contrast, primordial enzymes likely possessed substrate ambiguity (the ability to recognize a variety of molecules) and catalytic promiscuity (the ability to catalyze different chemical reactions with different substrates) *(2)*. These features are conducive for generating new enzymes with specialized functions through gene duplication and specialization *(2-4)*. In an organism under environmental stress, which potentially requires new reactions to survive, specificity of activity can be selected by evolution. Low levels of promiscuous activities may persist in a specialized enzyme if they are not detrimental to the organism, and therefore there is no selective pressure for their elimination *(5)*. Enzyme promiscuity can

also complement gene deletions in metabolic pathways, as in the case of *phn* operon knockouts in *E. coli* that recruit the phosphite-dependent hydrogenase activity of alkaline phosphatase *(6)*, or *E. coli* auxotrophs that utilize promiscuous sugar kinases encoded by cryptic genes to rescue glucokinase deficient mutants *(7)*. Thus, it has been suggested that promiscuity is an innate characteristic of all enzymes rather than an anomaly, and that there is a multitude of promiscuous activities extant in living cells merely waiting to be discovered.

In Chapter 2, the crystal structures of isopentenyl phosphate kinase from THA and MTH in ternary complexes with substrates and products were presented and became the newest structure in the AAK family of enzymes, which also includes NAGK, AK, CK, G5K, UMPK and FomA. In this family, IPK and the bacterial resistance protein fosfomycin kinase FomA in *Streptomyces* have very high structural similarity despite significant sequence divergence (22-25% identity). Moreover, the two enzymes possess the catalytic lysine triad and an active site histidine residue not found in other members of the family *(8, 9)*. Thus, it was suggested that the *ipk* gene was the source of *fomA*, acquired by *Streptomyces* via horizontal gene transfer, and converted into a specialized fosfomycin kinase through mutation and selection *(8, 10)*. This scenario is consistent with acquisition of resistance in bacteria constantly exposed to antibiotics *(11)*. In the case of *Streptomyces*, fosfomycin kinase activity (Figure 3.1) was added to other resistance strategies in its arsenal, which includes fosfomycin inactivation by opening of the oxirane ring catalyzed by the metalloenzymes FosA, FosB and FosX *(12-15)* (Figure 3.2a and b).

In this chapter, the results of the investigation of the substrate promiscuity of IPK from *Thermoplasma acidophilum* (THA IPK) will be presented. THA IPK was tested for

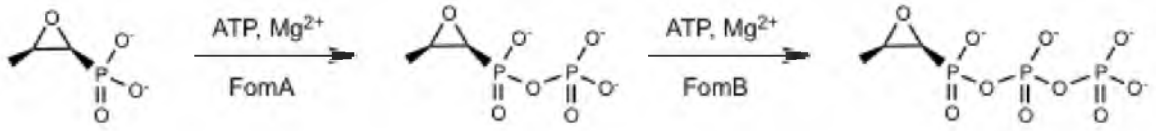


Figure 3.1 Fosfomycin inactivating reactions catalyzed by FomA and FomB found in the genomes of *Streptomyces* and *Pseudomonas*. FomA catalyzes the first phosphorylation step on fosfomycin, followed by FomB that adds the second phosphoryl group to make fosfomycin diphosphate. This product is unable to alkylate the active site cysteine residue of MurA, an enzyme critical for peptidoglycan synthesis.

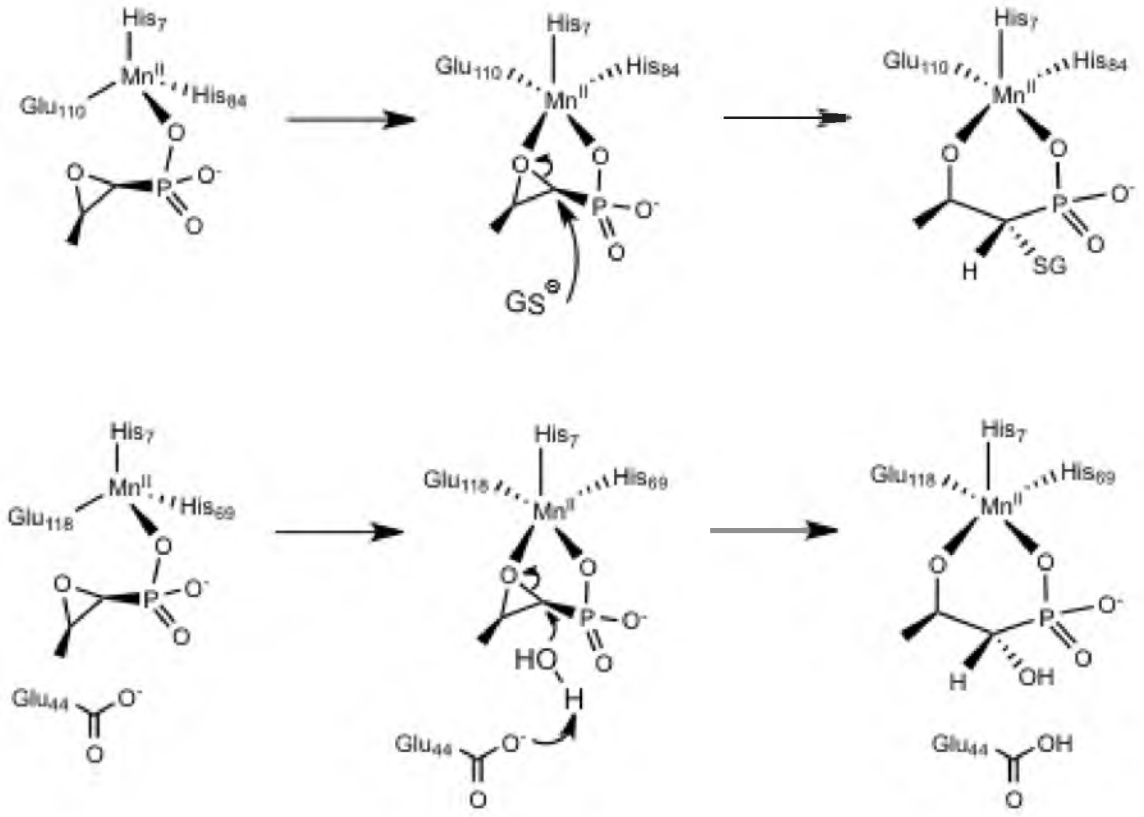


Figure 3.2 Mechanisms of fosfomycin ring opening reactions catalyzed by FosA and FosX. (Top panel) FosA uses a nucleophilic glutathione group in the active site to open the oxirane ring. FosB utilizes the same mechanism but uses a cysteine instead of glutathione. (Bottom panel) FosX is a hydratase that uses a water molecule to catalyze the ring opening reaction.

fosfomycin kinase activity and showed that the antibiotic is a promiscuous substrate for the enzyme, albeit with poorer binding affinity and a reduced rate of turnover. Moreover, fosfomycin is a competitive inhibitor of isopentenyl phosphate kinase activity at high concentrations. Using molecular dynamics simulations, two fosfomycin binding modes were discovered in which one is similar to that of IP in that it forms a stable complex with IPK, in addition to a separate binding mode that forms a dead-end complex.

Experimental Procedures

Materials. [^{32}P] ATP was purchased from American Radiolabeled Chemicals. Disodium salts of fosfomycin (>99%), phosphoenolpyruvate (PEP) and nicotinamide adenine dinucleotide (NADH) were purchased from Sigma-Aldrich. Bovine serum albumin was purchased from Invitrogen. Lactate dehydrogenase (LDH, rabbit muscle) and pyruvate kinase (PK, rabbit muscle) were purchased from Roche. The concentrations of all reagents containing phosphates were determined by phosphate analysis, and the precise concentration of stock IPK enzyme was measured using the BCA Protein Assay Kit from Pierce.

Synthesis of Isopentenyl Phosphate, IP (1). The procedure by Keller *et al.* was used to synthesize IP *(16)*. Briefly, a dry solution of 3.46 g (17.4 mmol) triethylammonium phosphate (TEAP) in 15 mL of acetonitrile was added dropwise to astirred mixture of alcohol (isoprenol, 3.25 mmol) in 5 mL of trichloroacetonitrile at room temperature under N_2, in three equal portions with a 5 min interval between additions. The mixture was then concentrated at reduced pressure, giving a thick, dark orange oil. The resulting oil was chromatographed on silica gel using 2-

propanol/NH_4OH/H_2O (6:3:1 v/v/v) and fractions containing pure products were combined. The solvent was removed by lyophilization, resulting in a white solid. Detection of IP was performed by thin layer chromatography using the same mobile phase and stained with *p*-anisaldehyde, $R_f = 0.50$; ^1H NMR (D_2O) δ 1.64 (s, 3H), 2.22 (t, 2H, $J = 6.6$ Hz), 3.79 (dt, 2H, $J = 6.9, 6.9$ Hz), 4.69 (s, 1H), 4.73 (s, 1H); ^{31}P NMR (D_2O) δ 1.13; HRMS (MALDI) calculated for $C_5H_{11}O_4P$ [M-H] 165.0322, found 165.0317.

Cloning and expression of THA IPK. The previous pET151/D-TOPO-THA IPK construct yielded proteins that were not trivial to purify. Due to the need for ultrapure proteins in kinetic experiments, a different construct based on the pET28b vector was constructed. The IPK gene in THA (TA0103) was amplified from genomic DNA (ATCC: 25905D) with PfuUltra high fidelity DNA polymerase (Stratagene) using the primers 5'-TGA TGA TAC TGA AGA TAG GCG GAA G-3' and 5' –AAA AGC CAA GCT TAT TAT CTT ATC ACC GTA CCT ATG AAT GAT TC -3'. The PCR product was digested with *Hind*III and ligated into pET28b (Novagen) prepared by *Nde*I digestion, Pfu polishing, and *Hind*III digestion to give the plasmid pET-THA. BL21 (DE3) cells (Novagen) were transformed with pET28b-THA IPK and were grown on agar plates with kanamycin. An overnight culture (10 mL) was grown using a single colony from the plates and was used to inoculate 6L of LB medium. The cultures were shaken at 37 °C until $OD_{600} \sim 0.5$ and induced with 1mM IPTG, and shaken at 30°C for another 6 h before centrifugation. Cell pellets were stored at -80 °C until needed.

Purification of THA IPK. Frozen cell pellets were thawed on ice and resuspended in lysis buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole). Lysozyme (1 mg/mL) was added to the suspension and incubated for 30 min before sonication on

ice. The crude lysate was centrifuged at 20,000 rpm at 4 °C for 25 min, and the resulting supernatant was incubated at 50 °C for 15 min followed by another centrifugation step. The supernatant was incubated with 5 mL Ni-NTA resin for 1 h at 4 °C with shaking (100 rpm). The column was washed with five column volumes each of lysis buffer and wash buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 50 mM imidazole), before elution with 25 mL of buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 250 mM imidazole) in 5mL fractions. SDS-PAGE was used to confirm the purity of the fractions, and those containing pure protein were combined, dialyzed at 4 °C against 20 mM Tris-HCl pH 8.0, containing 4 mM DTT, and stored in the same buffer containing 20% glycerol (v/v) as 100 μL aliquots at -80 °C. Prior to kinetic experiments, proteins were quantitated using the BCA protocol (Pierce).

Product turnover assay using γ-[^{32}P] ATP. The phosphorylation of fosfomycin by IPK was visualized by incubating the enzyme in assay buffer (100 mM HEPES pH 7.5 containing 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mg/mL BSA, γ-[^{32}P] ATP and varying concentrations of fosfomycin at 37 °C for 10 min. Each reaction was quenched with 113 uL of methanol/750 mM EDTA (100:13 v/v). Samples (5 μL) were spotted on silica plates and developed with $CHCl_3$/pyridine/formic acid/ H_2O (30:70:16:10 v/v/v/v). The TLC plate was imaged for 24 h using a storage phosphor autoradiography cassette and visualized using a Typhoon 8600 variable mode imager (GE Healthcare). The same procedure was used in comparing the radioactive product of IPK with that of *S. wedmorensis* FomA, kindly provided by Dr. Kate Slessor. The mixtures each contained 10 mM of fosfomycin and 10 uM enzyme, and incubated at 37 °C for 2 h.

MS/MS fragmentation of fosfomycin phosphate product. In order to confirm the presence of the fosfomycin phosphate product by IPK, we performed an MS/MS fragmentation experiment on the ammonium form of fosfomycin phosphate (mass = 236) produced by IPK and FomA. The said products were obtained by incubating 20 μM of each enzyme in assay buffer (100 mM HEPES pH 7.5, 1 mg/mL BSA, 10 mM β-mercaptoethanol, 10 mM $MgCl_2$) containing 25 mM fosfomycin at 37 °C for 2 hr, in a total volume of 1 mL. The enzymes were removed by centrifugation using a 10, 000 MWCO Centricon, and the filtrate was flash frozen in liquid N_2 and lyophilized overnight in 1.5 mL microfuge tubes. The resulting slurry was dissolved in 100 μL of 25 mM NH_4HCO_3 and submitted for MS/MS fragmentation.

Measurement of the activities of coupling enzymes. The activities of coupling enzymes were determined by measuring the change in absorbance of NADH at 339 nm. Different concentrations of lactate dehydrogenase (LDH) were mixed in the assay buffer (100 mM HEPES containing 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mg/mL BSA, 120 μM pyruvate and 150 μM NADH) at 37 ^{o}C. For pyruvate kinase, different concentrations of the coupling enzyme were mixed in assay buffer (100 mM HEPES containing 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mg/mL BSA, 1 mM PEP, 4 mM ADP, 150 μM NADH and LDH) at 37 ^{o}C. The enzymatic rates in AU/s where converted to specific activity units (U/mL) by using the NADH extinction coefficient $\varepsilon = 6.22\ mM^{-1}\ cm^{-1}$.

Steady state kinetics of the IPK-fosfomycin reaction. The protocol for fluorescent assays was based on the procedure by Pilloff *et al.* with slight modifications *(17)*. To initiate the reaction, IPK was added to the assay buffer (100 mM HEPES pH 7.5

containing 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mg/mL BSA and ATP) including appropriate amounts of coupling enzymes and fosfomycin (or the native substrate IP for positive control reactions) in a final volume of 200 µL. The reaction was monitored at 37 °C for 600 s by observing the change in fluorescence (λ_{ex} = 340 nm, λ_{em} = 460 nm) (FluoroMax, Jobin Yvon Horiba). Background rates were measured at different concentrations of fosfomycin in the absence of the enzyme and were averaged. The initial rates were measured from the linear portion of the curve (< 15 % consumption of the concentration-limiting substrate). The kinetic constants were determined by fitting the matrices of initial rates to equation 1 using Grafit 5 (Erithacus Software) *(18)*:

$$\nu = V_{max}[A][B] / K_d^A K_M^B + K_M^B[A] + K_M^A[B] + [A][B] \qquad (1)$$

where A and B are fosfomycin and ATP, V_{max} is the maximum rate, K_M is the Michaelis-Menten constant and K_d is the enzyme dissociation constant. The lag time τ of the coupled assay is given by the equation

$$\tau = K_M^{ADP}/V^{PK} + K_M^{Pyr}/V^{LDH} \qquad (2)$$

where K_M^{ADP} = 0.3 mM according to Sigma and K_M^{Pyr} = 0.164 mM according to Zewe and Fromm *(19)* and V is the activity of the enzyme in U/mL. The lag time τ of the reaction was then chosen to be less than 10 s, which ensured that the system reached 99 % of the steady state rate of ADP production in 45 s according to the equation

$$dP/dt = \nu_0 (1 - e^{-t/\tau}) \quad (3)$$

where ν_0 is the steady state rate and t and τ are in min.

Competitive inhibition of IP kinase activity by fosfomycin. Different concentrations of IP and fosfomycin were added to the assay buffer (100 mM HEPES pH 7.5 containing 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 1 mg/mL BSA). The reaction was initiated by mixing a solution of IPK in the same buffer with the first solution followed by incubation at 37 ^{o}C for 600 s. Initial rates were measured as described for the fluorescence assay, and the inhibition constant K_i was determined by fitting the initial rate matrices to the equation for competitive inhibition:

$$\nu = V_{max} [S] / [S] + K_M(1 + [I]/K_i) \quad (4)$$

where S is the substrate IP and I is the competitive inhibitor fosfomycin.

Molecular dynamics simulations of the IPK•fosfomycin•MgATP complex. Molecular dynamics simulations were performed using GROMACS *(20)* with CHARMM27 *(21)* force field for proteins and small molecules *(22)* and TIP3P *(23)* for water. The initial structure of the kinase is derived from the crystal structure of IPK *(8)*, while the specific binding pose of ligands in the binding pocket of IPK• fosfomycin•MgATP was derived from the superposition of structures 3LKK (IPK) and 3D41(FomA). The missing loop and αF helix of IPK (residues 189-201) were modeled based on sequence and structure alignment with NAGK (1GS5) where the corresponding

secondary structures in the latter most approximate the length of these missing regions (Figure 3.3a and b). The geometry of the modeled secondary structures was validated using Molprobity *(24)*. After an initial rapid energy minimization, the system was equilibrated by performing a position-restrained NVT simulation for 100 ps followed by an NPT run for another 100 ps. We then performed an unconstrained production run for 30 ns to validate the initially hypothesized poses of fosfomycin and MgATP. Long-range electrostatics was treated with Particle Mesh Ewald *(25)* with grid spacing of 12 Å and a cutoff of 10 Å. We used the modified Berendsen *(26)* and Parrinello-Rahman *(27)* algorithms for temperature and pressure couplings, respectively. The same experiments were performed on the native IPK•IP•MgATP complex as a control.

Site-directed mutagenesis of THA IPK. To create the Lys14Ala and Lys204Ala mutants, the forward primers 5'-GGC GGA AGC GTG ATC ACC GAT GCA TCC GCT TAC CG-3' and 5'-GTT ACC GGC GGT ATA GGG GCG AAG TTC GAA TCC ATG GT-3' were designed. PCR reactions were carried out following the protocol in the Quikchange Site-Directed Mutagenesis Kit manual (Stratagene)

Results

Detection of promiscuity. Autoradiography studies using γ-[^{32}P] ATP showed that fosfomycin is an alternate substrate for IPK. The formation of the product was accompanied by a concomitant decrease in γ-[^{32}P] ATP when the concentration of fosfomycin was in the high μM to mM range (Figure 3.4a). In contrast, K_M^{IP} = 4.4 μM for THA IPK. The product of the promiscuous reaction, assigned as fosfomycin phosphate, is consistent with the intense spot with R_f = 0.14 relative to IPP, R_f = 0.37.

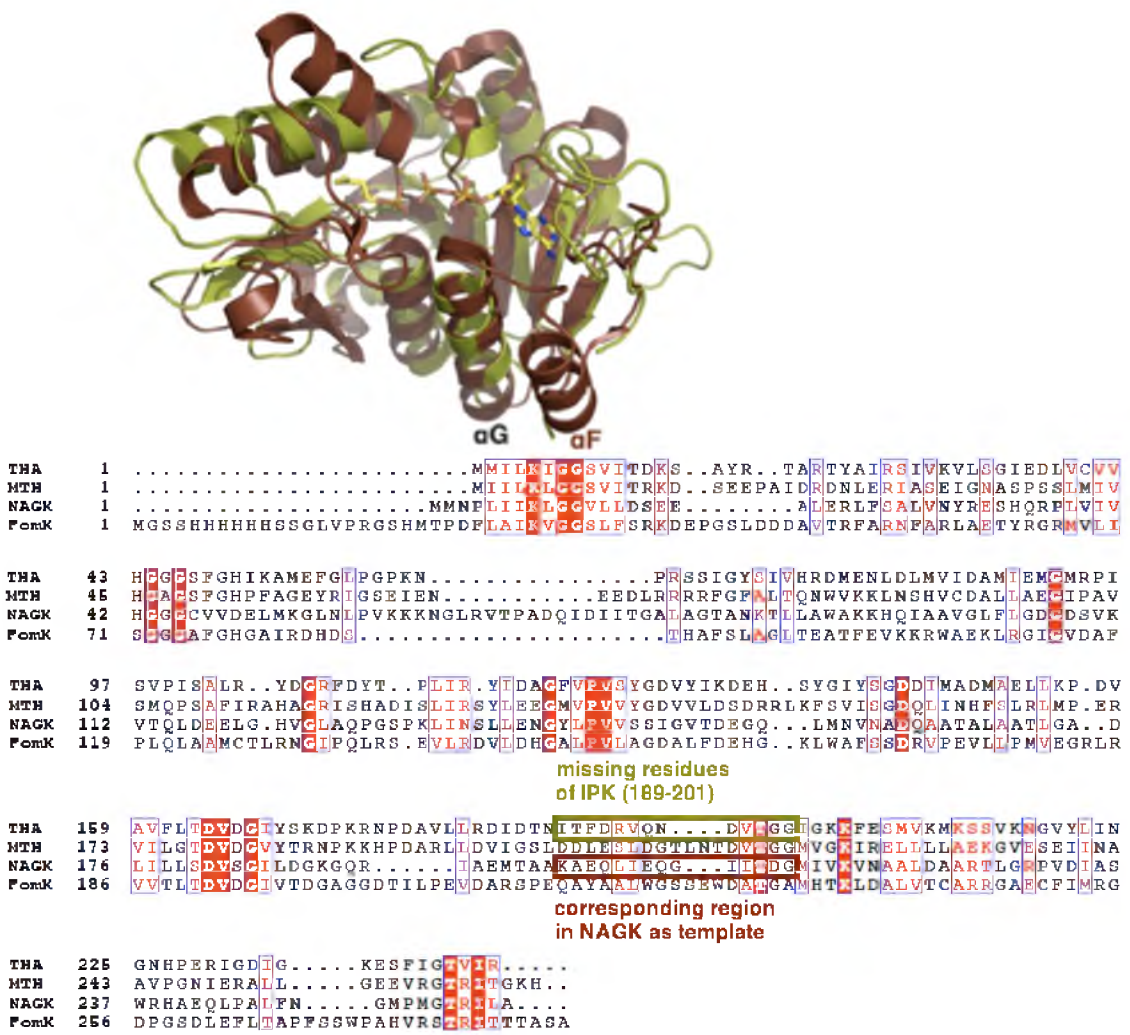


Figure 3.3 Sequence and structure alignment of IPK and NAGK enzymes. (Top panel) Superimposed IPK (green) and NAGK (red) structures showing the missing αF helix and the loop that connects it to the αG helix. (Bottom panel) Sequence alignment of homologous enzymes showing the stretch of sequence in NAGK used to model the missing structures in IPK.

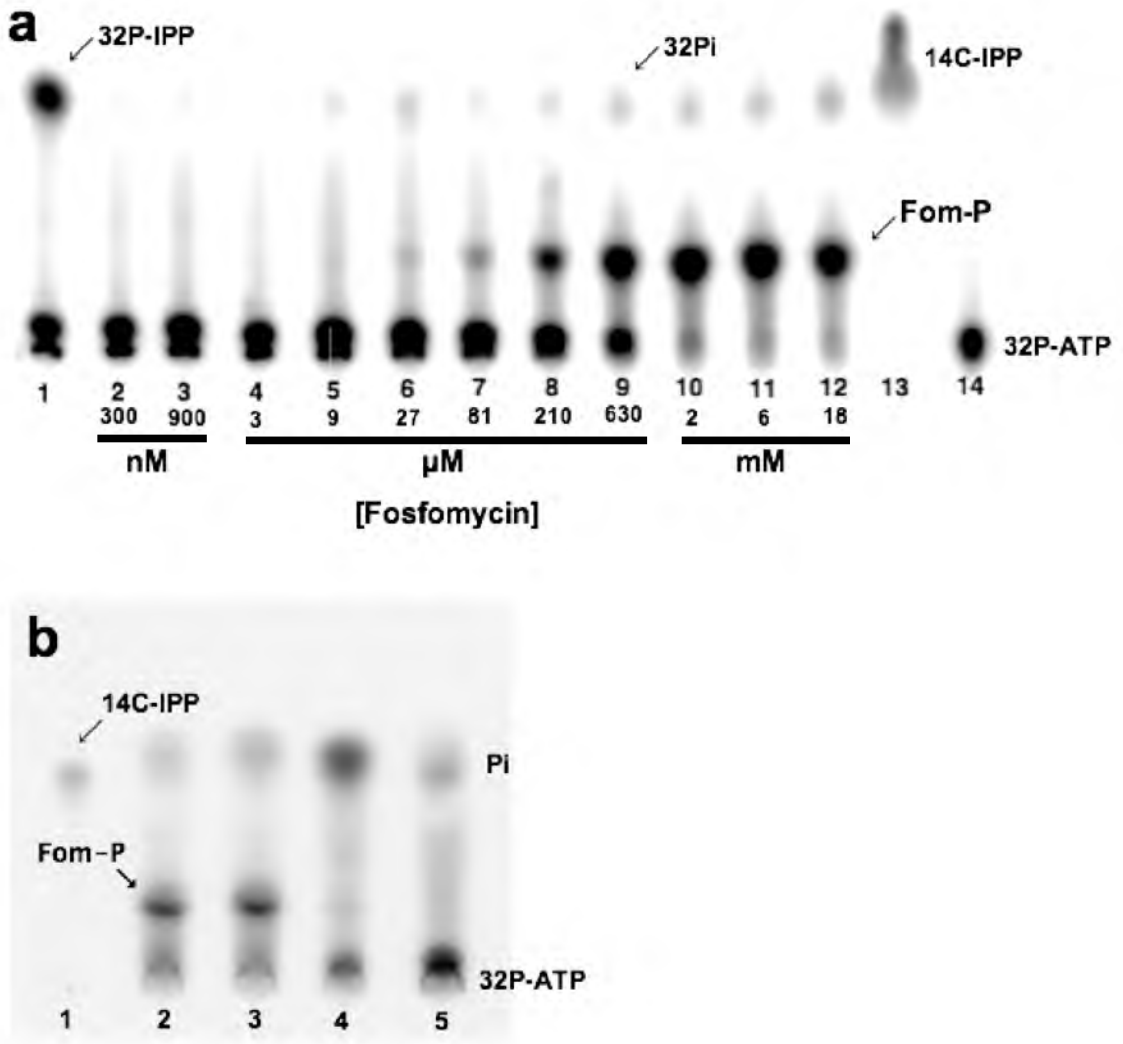


Figure 3.4 Autoradiogram showing that IPK can phosphorylate fosfomycin in the presence of ATP. (a) The product of the native reaction, β-[^{32}P] IPP, and the byproduct of the intrinsic ATPase activity, ^{32}Pi , migrate with almost similar R_fs. This ATPase activity is independent of fosfomycin concentration. Increasing phosphorylation of fosfomycin is observed in the μM to mM range. All lanes except lane 13 contained γ-[^{32}P] ATP and 200 nM IPK. (1) 3.5 μM IP and IPK, (2) 300 nM fosfomycin, (3) 900 nM, (4) 3 μM, (5) 9 μM, (6) 27 μM, (7) 81 μM, (8) 210 μM, (9) 630 μM, (10) 2 mM, (11) 6 mM, (12) 18 mM, (13) [^{14}C] IPP, (14) γ-[^{32}P] ATP and IPK. (b) Comparison of the products of *S. wedmorensis* FomA and IPK. (1) [^{14}C] IPP, (2) FomA and fosfomycin, (3) IPK and fosfomycin, (4) FomA without fosfomycin, (5) Fosfomycin without enzyme.

This spot did not appear in control reactions containing only IP or γ-[^{32}P] ATP and IPK. In addition, incubation of FomA and IPK showed product spots with the same R_fs as above (Figure 3.4b). These product spots were not found in incubations containing either enzymes or fosfomycin substrate. MS/MS fragmentation of the ammonium form of fosfomycin monophosphate (mass = 236) coming from incubates with IPK and FomA resulted in identical fragmentation patterns. Positive ion species corresponding to $[C_3H_5O]^+$ (mass = 57), $[C_3H_6O_3P]^+$ (mass = 121), $[H_4O_4P]^+$ (mass = 99), and $[C_3H_8O_4P]^+$ (mass = 139) were found in the fragmentations of fosfomycin monophosphate produced by either IPK or FomA. The first two cations, $[C_3H_5O]^+$ and $[C_3H_6O_3P]^+$ were also observed in the fragmentation of a fosfomycin standard (see Appendix).

In ^{31}P NMR assays, resonances corresponding to a phosphonate-phosphate moiety were not detected owing to the slow reaction and an apparently unfavorable equilibrium between fosfomycin-ATP and fosfomycin phosphate-ADP. We initially attempted longer incubation times without success. The autoradiogram shows a byproduct of the kinase reaction whose R_f is similar to that of IPP (Figure 3.4a), which is more pronounced in the absence of fosfomycin (Figure 3.4b, lane 4). Chen and Poulter *(10)* had previously shown that this by-product is $^{32}P_i$, resulting from the combined effects of γ-[^{32}P] ATP degradation and the intrinsic ATPase activity of IPK. ATPase activity has been observed in many kinases and is commonly negligible compared to native kinase activities *(28)*, although in some cases is sufficiently large to interfere with their precise measurements *(29)*. The ATPase activity of THA IPK was found to be sufficiently small (~ 200 times slower) that it did not interfere with measurements of the promiscuous fosfomycin kinase activity.

Kinetic constants of the promiscuous reaction. The ability of IPK to phosphorylate a variety of other phosphorylated small molecule substrates was recently reported by Chen and Poulter *(10)* and Dellas and Noel *(9)*. This promiscuity can now be extended to a small molecule bearing a phosphonate group. Initially, a series of fosfomycin concentrations ranging from 750 μM to 90 mM was used to measure the apparent kinetic constants of the reaction at saturating ATP concentration (Figure 3.5). These experiments showed that IPK could be saturated with fosfomycin and allowed us to identify a suitable range consisting of four different fosfomycin and ATP concentrations. Initial rate matrices were fitted to the equation for bisubstrate reactions (Figure 3.6a) and were transformed to a Lineweaver-Burk plot (Figure 3.6b). The fosfomycin kinase reaction proceeds by the same sequential mechanism seen for IP and ATP. The kinetic constants for the promiscuous and native activities of THA IPK are summarized in Table 3.1.

Fosfomycin is a competitive inhibitor of IPK. The k_{cat} for the promiscuous activity of IPK is at least two orders of magnitude smaller than that of the native kinase activity (Table 3.1). Moreover, superposition of the THA IPK and FomA ternary complexes aligned the respective IP and fosfomycin substrates (Figure 3.7). These kinetic and structural data indicate that fosfomycin can act as a competitive inhibitor of the faster IP kinase reaction. Kinetic studies with varying concentrations of fosfomycin were performed. The best fit was obtained with the equation for competitive inhibition

$$\nu = V_{max}[S] / [S] + K_M(1 + [I]/K_i) \qquad (1)$$

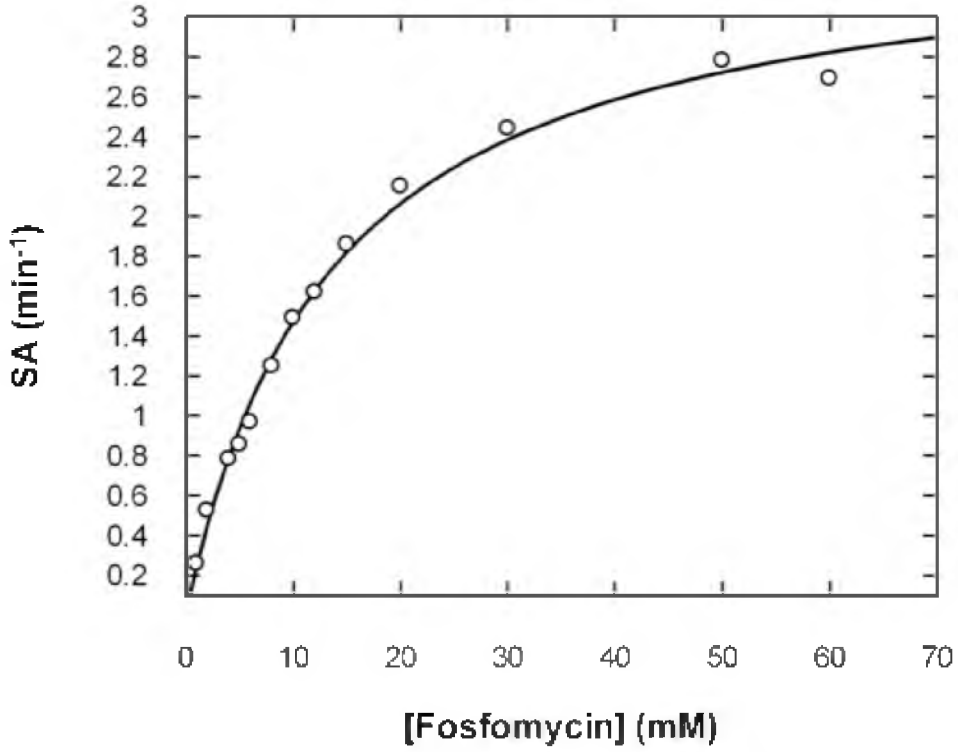


Figure 3.5 Michaelis-Menten saturation curve for IPK using fosfomycin as a substrate. Maximum specific activity is reached at high mM concentrations of fosfomycin.

Figure 3.6 Fosfomycin is a promiscuous substrate of IPK. (a) Michaelis Menten curve showing saturation of IPK at high mM concentrations of fosfomycin. A high K_M (15.1 mM) indicates weak binding of fosfomycin to the IP binding site. (b) Lineweaver-Burk transformation of (a), indicating a sequential binding mechanism and formation a ternary complex of IPK•fosfomycin•ATP. (○) 2 μM ATP, (●) 4 μM ATP, (□) 6 μM ATP, (■) 8 μM ATP.

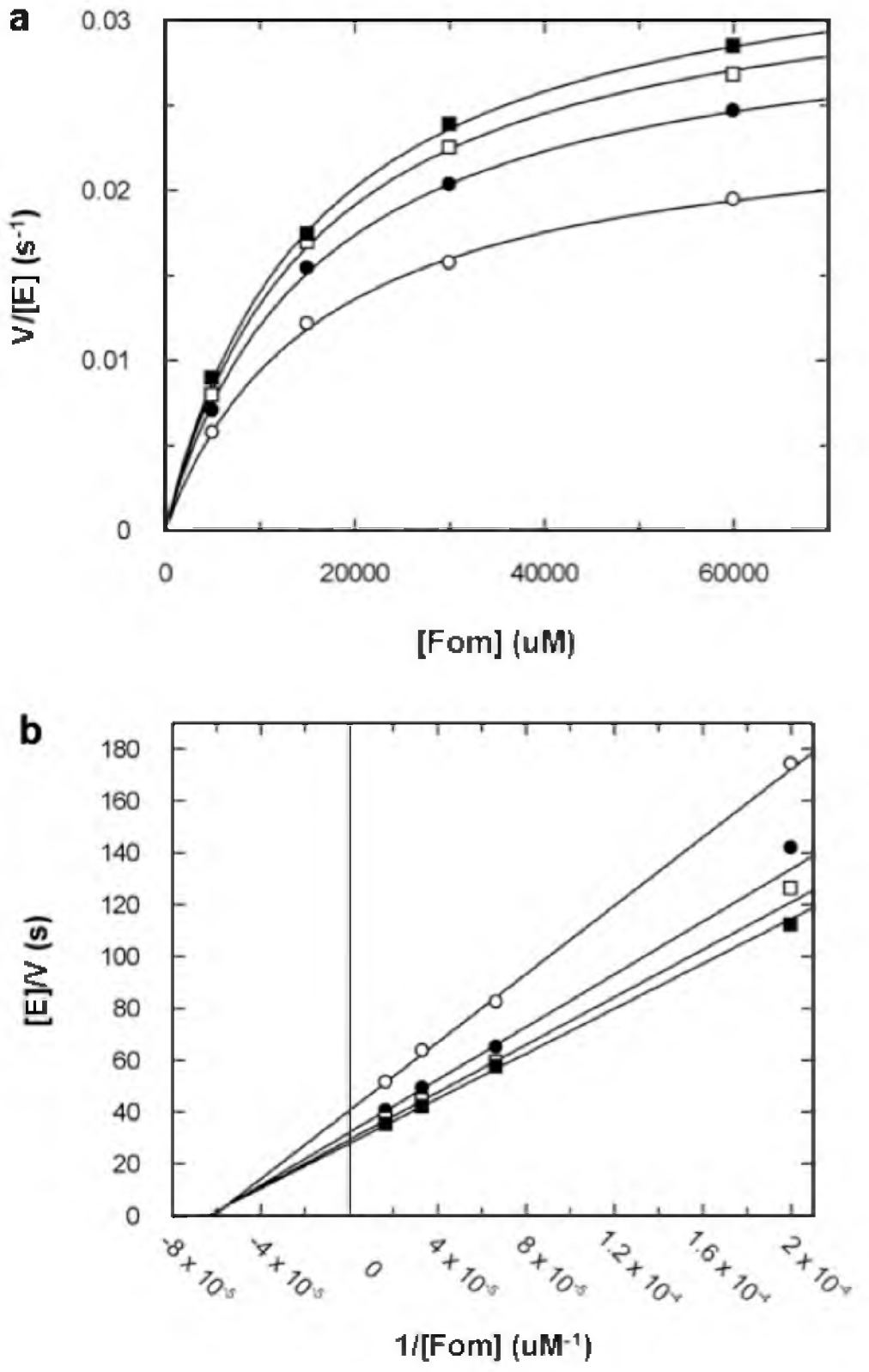


Figure 3.6 continued

Table 3.1: Steady state kinetic constants for the native and promiscuous activities of THA IPK.

	IP Wild-type IPK	**Fosfomycin Wild-type IPK**	**IP K204A**	**Fosfomycin K204A**
k_{cat} (s^{-1})	8.0 (± 0.2)	4 (± 0.1) × 10^{-2}	7.7 (± 0.2)	2.7 (± 0.1) × 10^{-2}
K_M(μM)	4.4 (± 0.5)	1.5 (± 0.1) × 10^{4}	2.8 (± 0.3)	1.1 (± 0.1) × 10^{4}
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.8 × 10^{6}	2.6	2.8 × 10^{6}	2.5
K_M^{ATP} (μM)	6.0 (± 0.5)	1.4 (± 0.1)	6.9 (± 0.6)	1.2 (± 0.2)
K_d (μM)	4.6 (± 1.5)	1.8 (± 0.3) × 10^{4}	4.6 (± 1.3)	1.5 (± 0.5) × 10^{4}

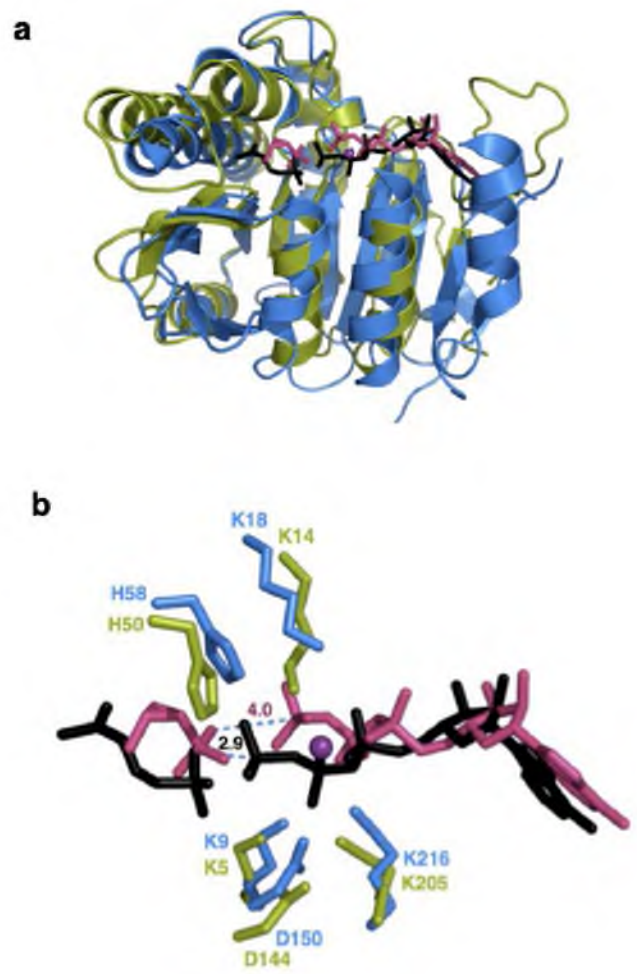


Figure 3.7 *T. acidophilum* IPK and *S. wedmorensis* FomA have high level of structural similarity. (a) Superimposed crystal structures of IPK (3LKK, green) and FomA (3QUO, blue) have RMSD of 2.6 Å over 224 amino acid residues. The substrates of IPK (IP and ATP, black) and FomA (fosfomycin and ATP, pink) also align well in their respective active sites. The divalent metal Mg^{2+} is shown as a purple sphere. (b) Alignment of catalytic residues invariant in IPK and FomA, suggesting similar mechanisms of phosphoryl transfer. The FomA•fosfomycin•MgATP complex was modeled using the structures of the FomA•MgATP (3QUN) and FomA•fosfomycin•ATP (3QUO) complexes. Distances between the nucleophilic oxygen atoms of IP or fosfomycin and the electrophilic P atom being attacked are indicated.

where S is the substrate IP and I is the inhibitor fosfomycin. The Michaelis-Menten curves (Figure 3.8a) showed an apparent increase in K_M^{IP} with increasing inhibitor concentration, while at saturating concentrations of IP, the inhibitory effects of fosfomycin are overcome. The Lineweaver-Burk transformation of these curves showed an array of intersecting lines on the ordinate (Figure 3.8b), with k_{cat} = 7.5 s^{-1}, similar to the reported value in the literature (8.0 s^{-1}) and K_i = 3.6 ± 0.20 mM. Weak inhibition of THA IPK by fosfomycin is most likely due to the incompatibility of its polar epoxide group with the hydrophobic pocket tailored to the aliphatic tail of IP.

Binding poses of FomA in IPK. The kinetic data and the highly analogous structures of IPK and FomA strongly suggest that the native substrate IP and the promiscuous substrate fosfomycin adopt similar binding poses in IPK. To examine this issue, molecular dynamics simulations of the presumed pose of fosfomycin were performed using the superimposed structures of the IPK and FomA ternary complexes. To fix the MgATP ligand in its binding site during simulation, an accurately modeled structure for the missing αF-αG loop and αF helix of IPK was supplied to the starting structure based on the aligned sequences of IPK from *T. acidophilum* and *M. thermautotrophicus (8), S. wedmorensis* FomA *(30)* and *E. coli* NAGK *(24)* (Figure 3.3). These alignments indicate that the secondary structure in NAGK, which only differs from the sequence in THA IPK by a single amino acid, can be used to model the missing loop and αF helix in IPK. The system is stable and has an average Cα RMSD of 1.5 Å with respect to the starting structure (Figure 3.9a and b). The heavy atom RMSD of fosfomycin and ATP were also computed and showed two stable conformers of fosfomycin in its binding pocket (Figure 3.10a). The normalized RMSD distributions

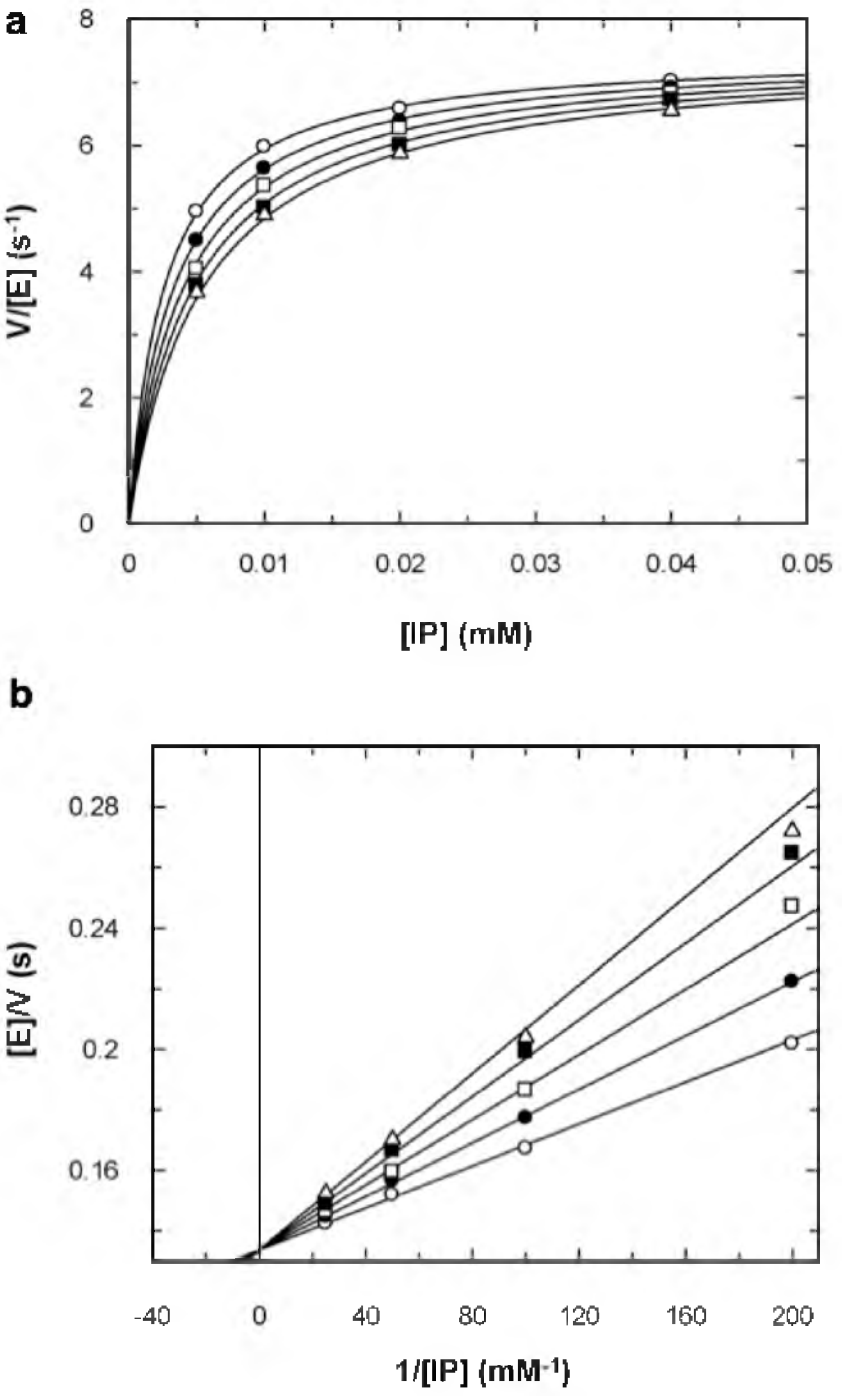


Figure 3.8 Fosfomycin is a competitive inhibitor of IPK. (a) Michaelis Menten curve showing an increase in K_M^{IP} with increasing fosfomycin concentration. (b) Lineweaver-Burk transformation of (a) showing a family of lines intersecting at the y-axis, indicating a constant V_{max}. (○) 0 mM fosfomycin, (●) 1 mM fosfomycin, (□) 2 mM fosfomycin, (■) 3 mM fosfomycin, (△) 4 mM fosfomycin.

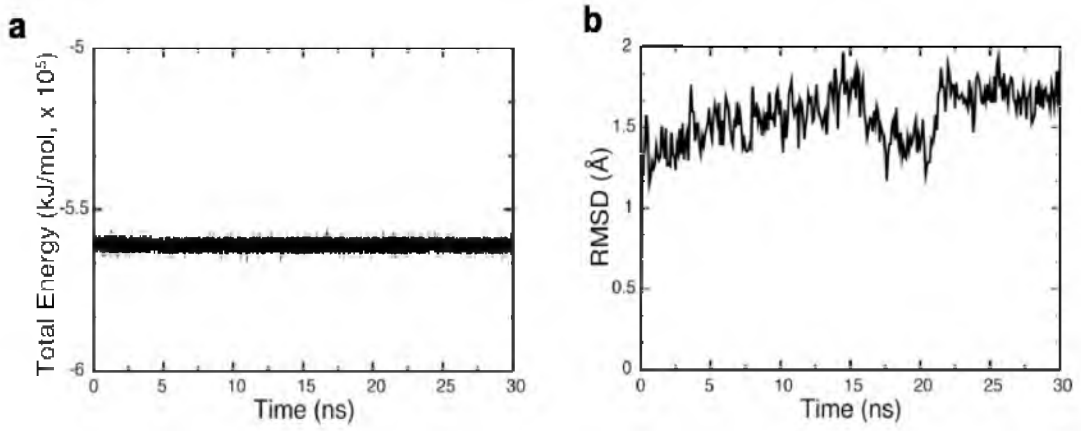
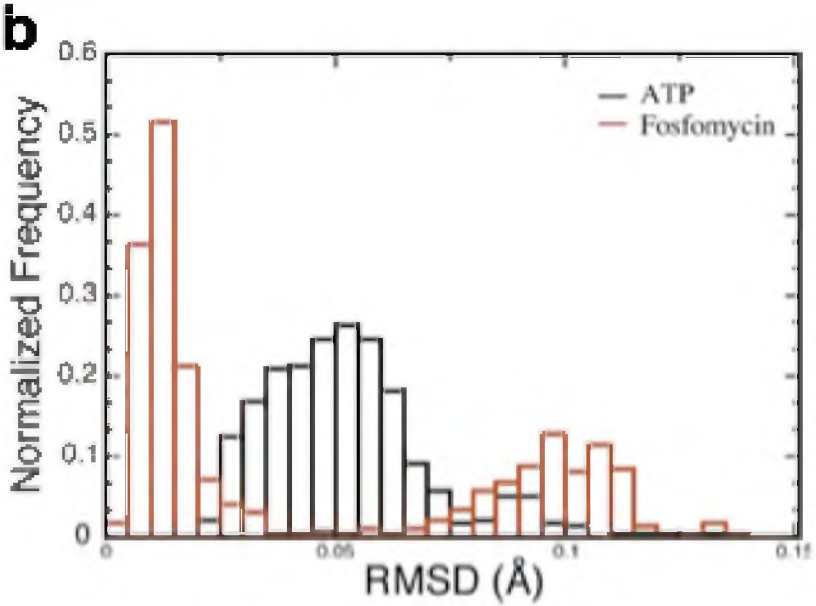
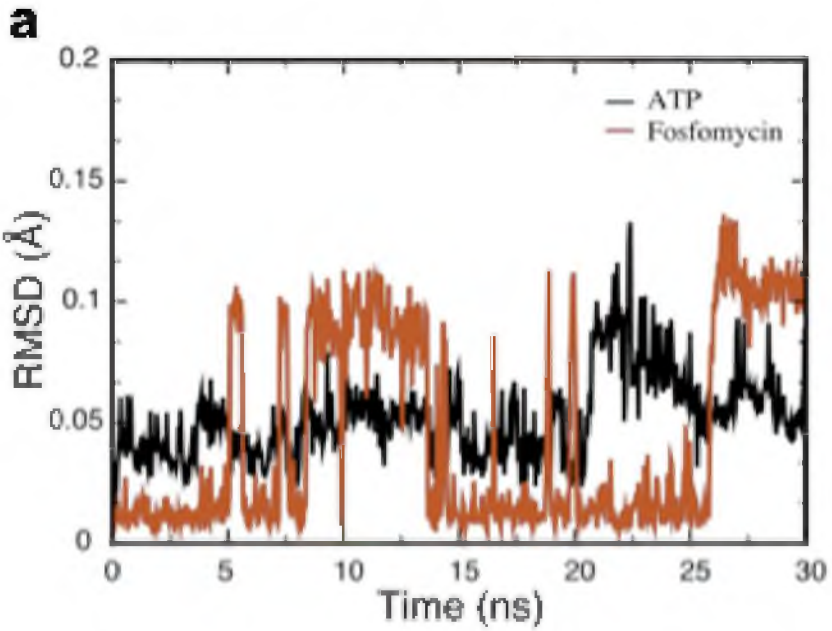


Figure 3.9 Energy (a) and RMSD time-series (b) showing that the modeled IPK•fosfomycin•MgATP complex is stable.

Figure 3.10 Ligand (fosfomycin and ATP, left) RMSD (a) and histogram (b) showing a two-state binding conformation for fosfomycin, while ATP is essentially fixed.


b
0.6
0.5
0.4
0.3
0.2
0.1
0
Normalized Frequency
ATP
Fosfomycin
0
0.05
0.1
RMSD (Å)


a
0.2
0.15
0.1
0.05
0
RMSD (Å)
ATP
Fosfomycin
0
5
10
15
20
25
30
Time (ns)

support the observed two-state conformation (Figure 3.10b). Closer inspection of these two conformations shows that fosfomycin binds to the N-terminal lobe of THA IPK that contains the IP binding site (Figure 3.11).

In both complexes, fosfomycin can obstruct the IP binding site, supporting the observed competitive mode of inhibition. Moreover, in both complexes the MgATP conformation is essentially fixed while fosfomycin is translated by a few Å (3.9 Å in the first pose and 4.8 Å in the other pose) away from the location of IP in the binding site, presumably to reduce unfavorable interactions between the polar epoxide and the hydrophobic binding pocket (Figure 3.11). In the first binding pose (Figure 3.12a), the oxygen of the epoxide is oriented toward the solvent region while the methyl group packs into the hydrophobic pocket. The catalytic residue His50 in the αB helix forms a hydrogen bond with the phosphonate group, analogous to its interaction with IP. The $N\varepsilon_1$ nitrogen atom is the hydrogen bond donor in this binding pose, instead of the $N\varepsilon_2$ nitrogen atom observed in the crystal structure *(8)*. Nevertheless, this interaction positions fosfomycin for attack on ATP, where the γ-phosphate is 5.1 Å away. This is substantially longer than the 2.9 Å distance between the IP phosphate and γ phosphate of ATP in the IPK ternary complex (3LKK) and likely contributes to the low efficiency of the promiscuous reaction. However, the distance between the phosphonate group and the γ phosphate of ATP in the structure of the FomA•fosfomycin•MgATP complex, modeled from coordinates of the FomA•MgATP (3QUN) and the FomA•fosfomycin•ATP (3QUO) complexes, is 4.0 Å (Figure 3.7) *(30)*. Thus, it is likely that this distance is a good approximation of the real phosphonate-γ-phosphate distance. Examination of this binding pose also reveals the active site Mg^{2+} is located 3.7 Å from a phosphonate

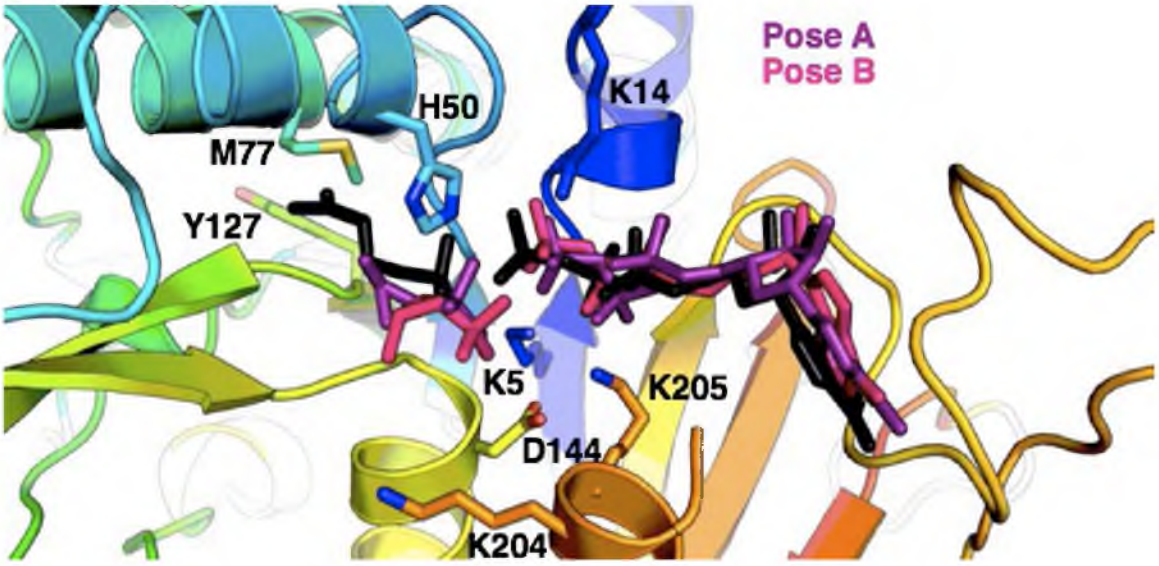


Figure 3.11 Superposition of the two observed binding poses (purple and pink) of fosfomycin to that of IP in its binding site in IPK. The two hydrophobic residues in the IP binding site, Met77 and Tyr127, are shown as sticks to emphasize the incompatibility of the polar fosfomycin epoxide ring to this site, resulting in the translation of fosfomycin away from the site in both poses. The catalytic residues of IPK that are conserved in FomA are shown as sticks. The ATP molecules are colored as the corresponding IP or FomA substrate.

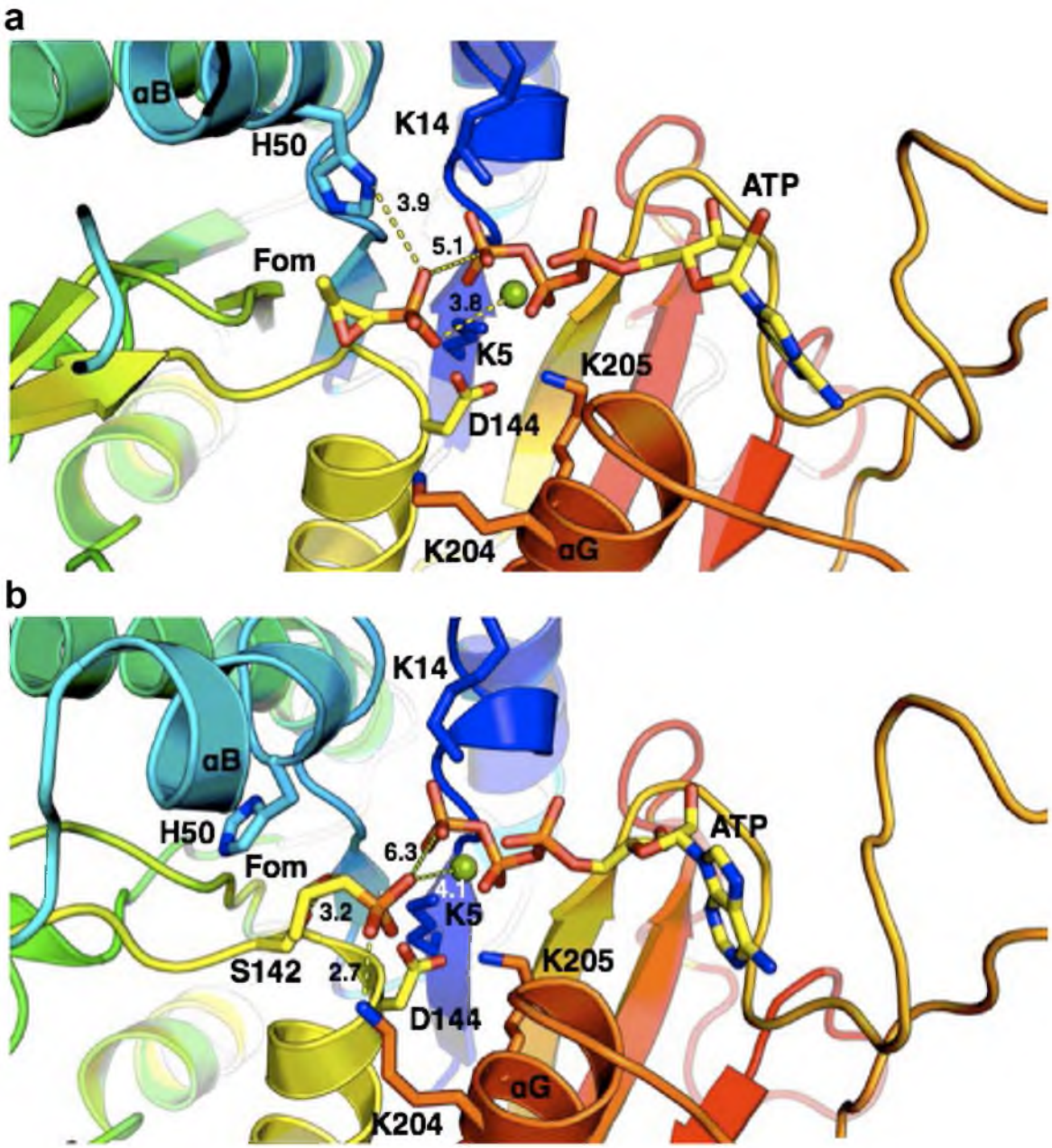


Figure 3.12. Binding poses of fosfomycin in the IP binding site of IPK. (a) Fosfomycin in the first orientation or pose. (b) Fosfomycin in the second orientation or pose.

oxygen, perhaps facilitating nucleophilic attack on ATP. This interaction as well as innate enzyme plasticity could help overcome the ~5 Å distance between reactive phosphonate and phosphate centers. The interactions of Lys5, Lys14, Lys205 and Asp144 observed in the IPK•IP•ATP complex *(8)* are essentially conserved in this binding pose.

In the second, less populated binding state, fosfomycin is removed even farther away from the precise IP binding site, eliminating hydrogen bonding with His50 and causing hydrogen bonds to form between the epoxide group and Ser142, and between the phosphonate group and Lys204 in the αG helix. In THA IPK, Ser142 binds the phosphate group of IP. This residue is conserved in FomA and orients fosfomycin for nucleophilic attack on ATP by hydrogen bonding with the epoxide O atom *(29)*. In addition, the phosphonate group is 6.3 Å away from the γ phosphate in ATP and 4.1 Å from Mg^{2+} in this second pose (Figure 3.12b).

Molecular dynamics simulation of IPK•IP•MgATP showed that the complex is stable (Figure 3.13) and revealed a single binding pose with IP and MgATP virtually unchanged from their respective poses in the crystal structure (Figure 3.14). The latter result suggests that the crystal structure closely approximates the structure of the complex in solution and also lends credence to the two observed binding poses of fosfomycin in the IPK active site.

The K204A mutant IPK has fosfomycin activity comparable to the wild type IPK. The recruitment of Lys204 to form a hydrogen bond to an oxygen atom of the phosphonate moiety appears in the simulation to "rescue" the His50-phosphonate hydrogen bonding interaction absent in the second binding complex. To validate the contribution of the second binding pose of fosfomycin and the resulting complex to

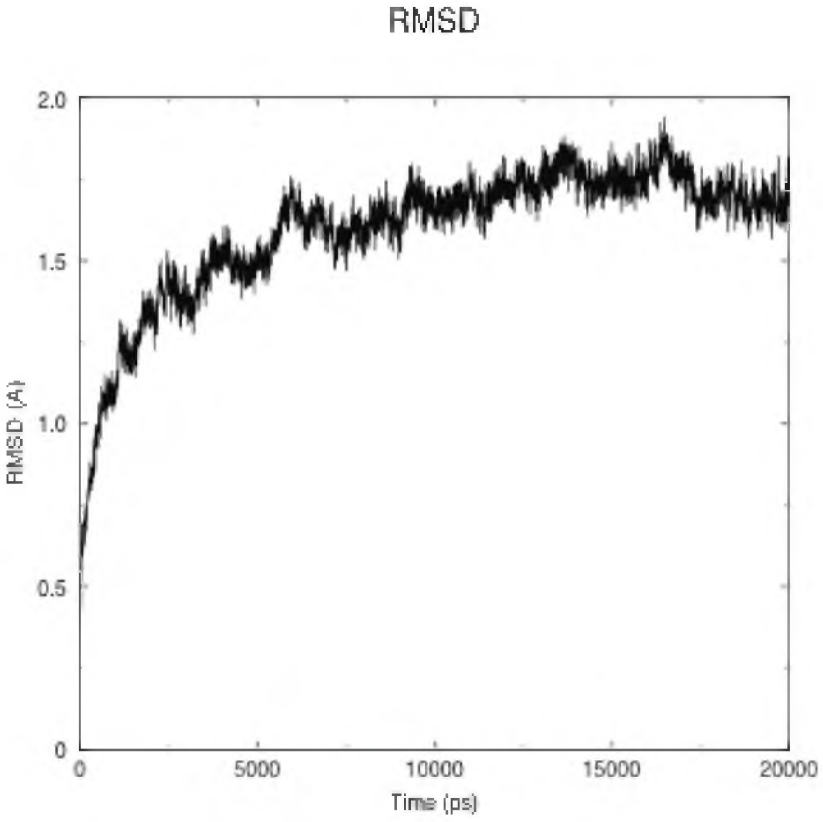


Figure 3.13 RMSD time-series showing that the modeled IPK•IP•MgATP complex is stable.

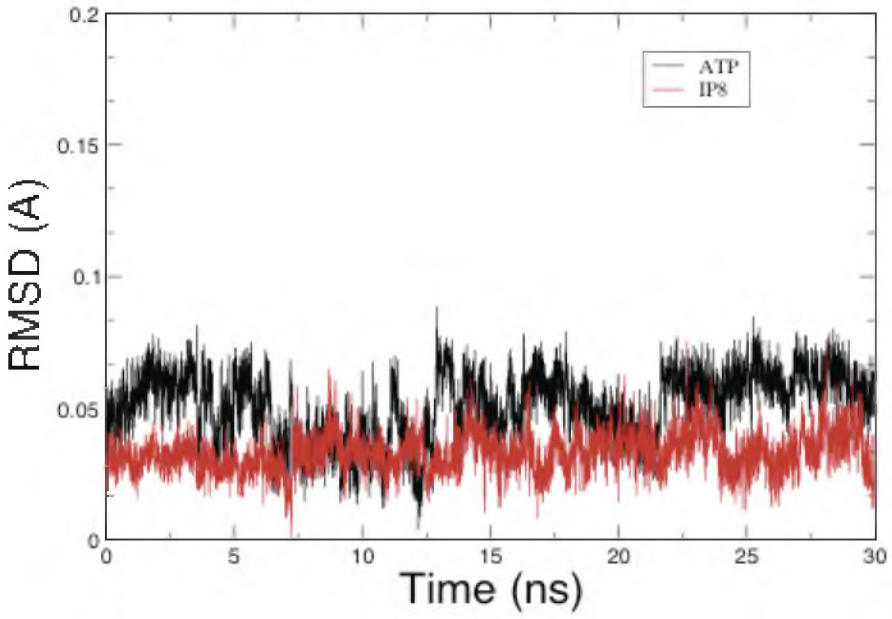


Figure 3.14 Ligand (IP and ATP) RMSD showing a single-state binding conformation for IP and ATP.

catalysis, a Lys204Ala mutant was purified and assayed for fosfomycin kinase activity. This mutant remained active and was saturated with fosfomycin at high mM concentrations like the native enzyme (Figure 3.15). There was only a 33% decrease in the turnover number k_{cat} as well as in the Michaelis constant, K_M, for fosfomycin, while the catalytic efficiency k_{cat}/K_M remained unchanged (Table 3.1, Column 5). Moreover, the mutation did not affect catalysis of the native reaction (Table 3.1, Column 4, Figure 3.16), as the kinetic parameters remained unchanged from those previously reported *(10)*.

Discussion

Fosfomycin is a secondary metabolite produced in several species of *Streptomyces (31)* and *Pseudomonas (32)* that is widely used clinically to treat cystitis *(33)* as well as methicillin- and vancomycin-resistant strains of *Staphylococcus aureus (34,35)*. Its bactericidal activity is based on the inhibition of MurA *(36)* by the irreversible alkylation of an active site cysteine residue, preventing the first step in peptidoglycan synthesis in which PEP is linked to the 3'-hydroxyl group of UDP-N-acetyl-glucosamine (Figure 3.17) *(37)*. Bacterial resistance to fosfomycin is achieved by mutations in MurA and by alterations in the glycerophosphate transporter *(38, 39)*. In addition, fosfomycin is inactivated by hydrolysis of the epoxide moiety by the thiol transferases FosA and FosB, and the homologous hydratase enzyme FosX (Figure 3.2a and b) *(12-15, 40)*. Two additional enzymes, FomA and FomB, alter the antibiotic by two consecutive phosphorylation steps resulting in a product unable to alkylate the relevant cysteine residue in MurA. The *fomA* and *fomB* genes are found in the fosfomycin

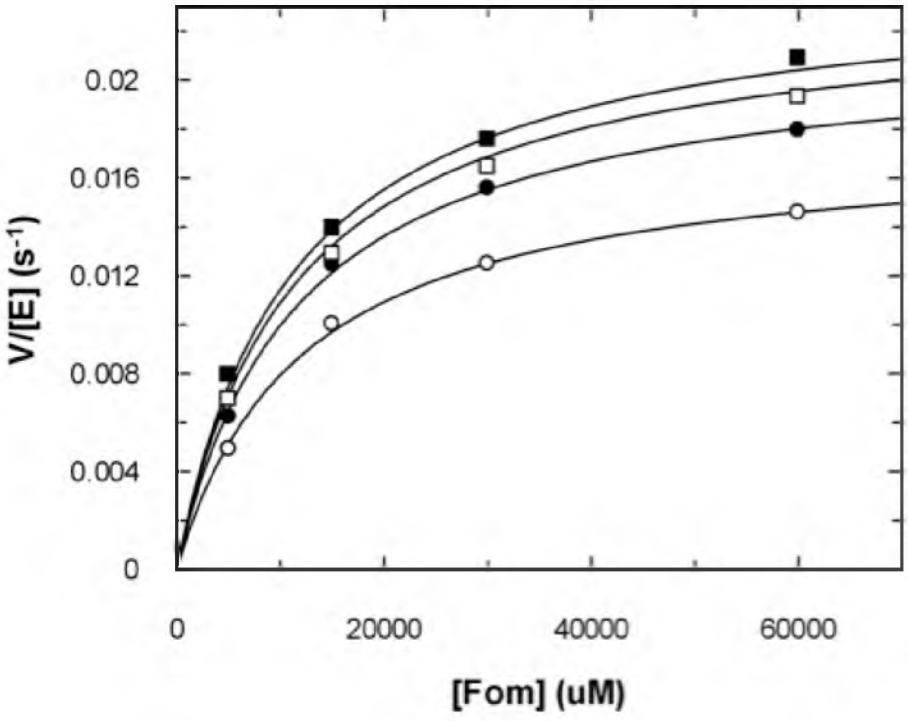


Figure 3.15 Michaelis-Menten curves from kinetic studies on the Lys204Ala IPK mutant with fosfomycin. (a) Kinetic measurements using fosfomycin as substrate showed that the mutant can catalyze the promiscuous reaction to the same extent as native THA IPK. (○) 2 μM ATP, (●) 4 μM ATP, (□) 6 μM ATP, (■) 8 μM ATP.

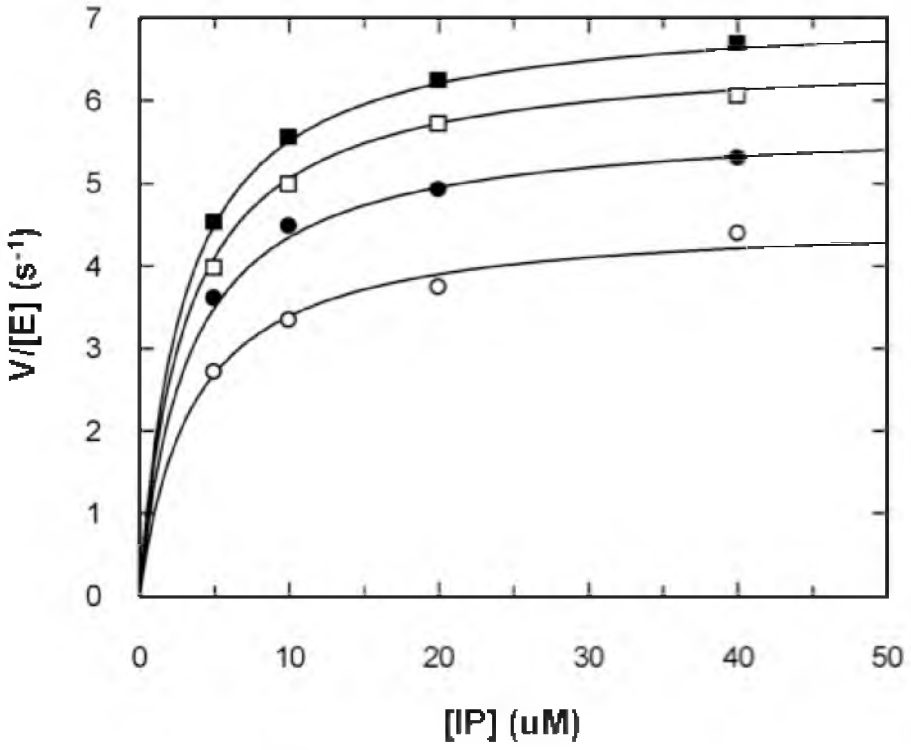


Figure 3.16 Michaelis-Menten curves showing that the Lys204Ala mutant can catalyze the native reaction (using IP as substrate) to the same extent as native THA IPK. (○) 10 μM ATP, (●) 20 μM ATP, (□) 40 μM ATP, (■) 80 μM ATP.

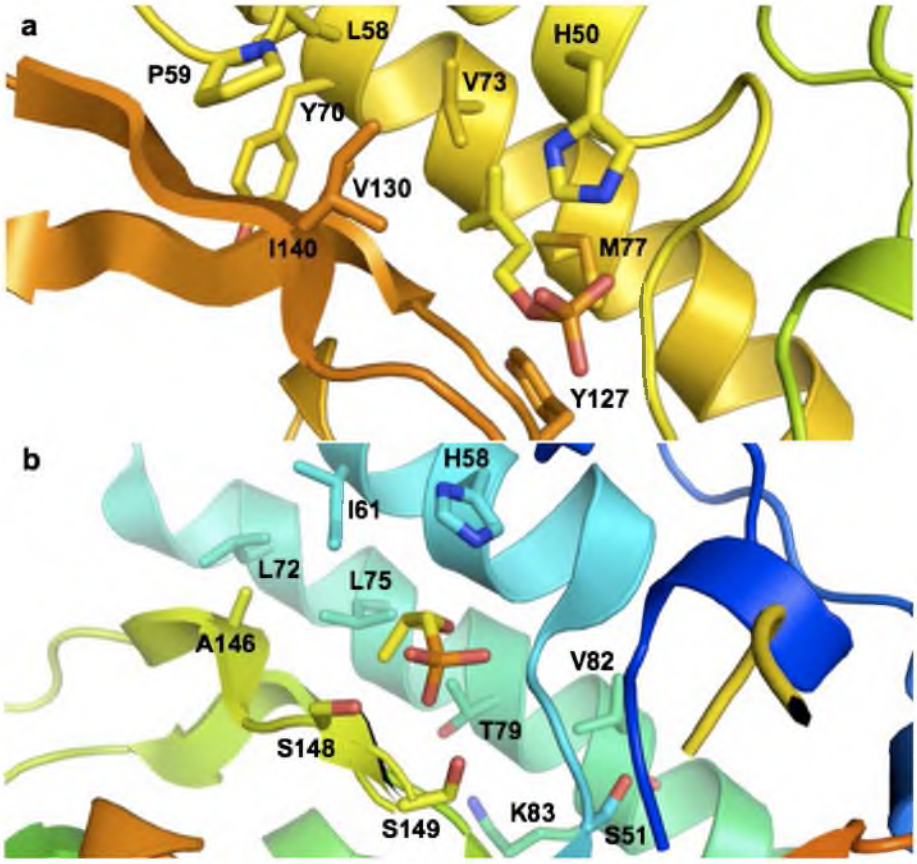


Figure 3.17 Comparison of the IPK and FomA binding sites. The hydrophobic side chain residues of both IPK and FomA are shown. The IP binding site of IPK is made of mainly hydrophobic residues, with Met77 and Tyr127 immediately below the aliphatic tail of IP. The fosfomycin binding site in FomA also contains some hydrophobic side chains aside from hydrophilic ones of Thr79, Lys83 and Ser51.

biosynthetic gene clusters of *Streptomyces* and *Pseudomonas* and are thought to confer resistance and facilitate transport in these hosts *(41, 42)*.

The phosphate subdivision of the AAK family of enzymes consists of IPK, FomA and uridine monophosphate kinase (UMPK) *(8, 9, 43)*. These three enzymes catalyze phosphoryl transfer from ATP via nucleophilic attack by an oxygen atom in a phosphate or phosphonate group in a reaction that proceeds through a pentavalent transition state intermediate *(8, 30, 43)*. Within this subdivision, IPK and FomA displayed significant structural similarity. Moreover, superposition of crystal structures for IPK•IP•ATP and the modeled structure of the FomA•fosfomycin•MgATP complex showed good alignment of their catalytic residues and respective substrates. The structural similarities between THA IPK and *S. wedmorensis* FomA suggest an evolutionary relationship between the two proteins, with an archaeal IPK as the likely ancestor. In that case, IPK might possess promiscuous activity toward the substrate for FomA, as is the case with many enzymes within a family *(8,9)*.

We measured the kinetic parameters for the promiscuous fosfomycin kinase activity of IPK by coupling the production of ADP to the oxidation of NADH, using pyruvate kinase and lactate dehydrogenase as coupling enzymes *(17)*. The turnover number, k_{cat} for the promiscuous reaction (0.04 s^{-1}) is 200-fold lower than that for the native activity (8.0 s^{-1}), while the Michaelis constant (K_M) for fosfomycin (15.1 mM) is more than three orders of magnitude greater than that for IP (4.4 μM). The catalytic efficiency for phosphorylation of fosfomycin by THA IPK (k_{cat}/K_M = 2.5 $M^{-1}s^{-1}$), while five orders of magnitude lower than for phosphorylation of IP, could still serve as starting point for evolution of a more potent fosfomycin phosphorylation enzyme. There are

notable examples in the literature of enhancement of secondary enzyme activities achieved by mutation of one or more residues in the active site. For instance, the *E. coli* RTEM β-lactamase increased its secondary cephalosporinase activity 1.4-fold and decreased its primary penicillase activity 3-fold upon placement of the bacterial strain under the selective pressure of cephalosporin C, and the observed changes in activity were caused by two mutations in the active site *(44)*. Subsequently, other mutations were found to cause more dramatic improvements in cephalosphorinase C activity, such as the Ala237Asn mutation observed by Healey *et al. (45)* and different mutant combinations observed by other groups *(46, 47)*.

The reduction in catalytic efficiency results primarily from the conformation of fosfomycin in the IP binding site relative to the native substrate. This site is laden with hydrophobic residues that cradle the aliphatic tail of IP and dictate the size of the isoprene moiety that can be accommodated in the active site *(8, 9)*. Fosfomycin, although smaller than IP, contains a polar epoxide group that appears to reduce binding to the hydrophobic pocket. A comparison of the active sites of FomA and IPK shows that the latter has a more hydrophobic region immediately below the aliphatic tail of IP, arising primarily from the side chains of Met77 and Thr127 (Figure 3. 17a) *(8)*. Similar residues are not found in the corresponding site of FomA (Figure 3.17b) *(10)*. Moreover, Ser149, which is crucial in binding fosfomycin in FomA, is replaced by a glycine residue in THA IPK and MTH IPK *(30,* Figure 3.3, lower panel*)*. Amino acids Gly53, His58 and Ser148 in FomA create a constellation of possible hydrogen bond donors to the epoxide group and phosphonate groups of fosfomycin. Although a similar hydrogen bonding network is possible in IPK *(8, 9)*, the immediate vicinity of the epoxide ring in FomA is not as

hydrophobic as the corresponding region in IPK (Figure 3.18a and b). In addition, the side chains of Thr79, Lys83 and Ser51 enhance the hydrophilicity of the fosfomycin binding site in FomA (Figure 3.17b). It was then not surprising that in both binding poses observed during simulation, fosfomycin was translated by a few Angstroms away from the IP binding site to relieve the unfavorable interactions between the hydrophobic pocket and the polar epoxide group, and in the first pose, to orient the epoxide group toward the solvent. These new positions still place fosfomycin in a blocking position in the IPK binding site, consistent with it being a competitive inhibitor of IP. Thus, the differences between the hydrophobicities of the active sites of IPK and FomA and the absence of a corresponding residue for Ser149 in THA IPK, explain the high K_M for the promiscuous substrate.

The low level of promiscuous activity of IPK can be explained by phosphorylation occurring in the first binding pose for fosfomycin, where the antibiotic forms a hydrogen bond with His50, the catalytic residue responsible for orienting the substrate for nucleophilic attack on ATP in the native reaction. Although the distance between the nucleophilic phosphonate group and the γ-phosphate group of ATP is longer than the phosphate group in the native reaction (Figure 3.7), it is similar to the native IPK•IP•ATP structure and can therefore be reasonably assigned as a productive complex. In addition, this binding complex engages the conserved Lys14 in the same interactions observed in the IPK•IP•ATP complex (Figure 3.12a). Lys14 binds the phosphate group of ATP in the IPK•IP•ATP complex and is positioned to stabilize the highly negatively charged transition state *(8)*. These roles have been confirmed biochemically, with the Lys14Ala mutation resulting in a two-fold increase in K_M^{ATP} and a 460-fold decrease in

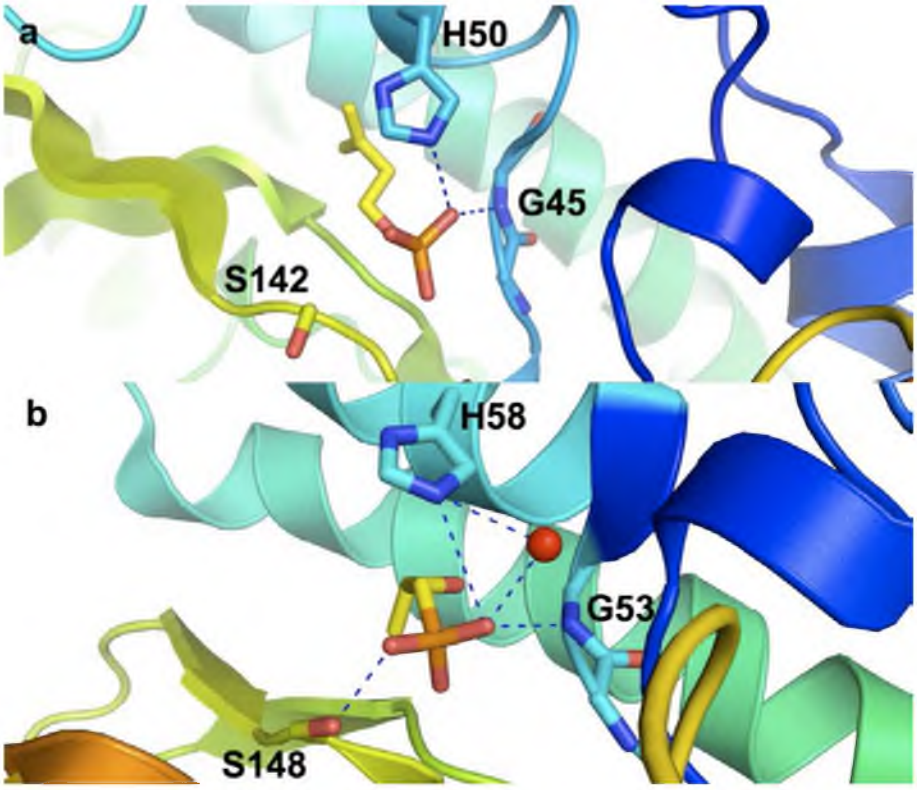


Figure 3.18 Side chains of residues in the IPK (a) and FomA (b) active sites that help orient IP and fosfomycin for nucleophilic attack by hydrogen bonding. Although Ser142 in IPK is in a different conformation in the crystal structure than the corresponding residue in FomA, it can undergo rearrangement to form a hydrogen bond as the corresponding serine residue in FomA (b).

k_{cat} compared to the native enzyme. The same mutation on the corresponding Lys18 residue in FomA resulted in an inactive enzyme *(30)*. The structure of the second complex suggests that it is not suitable for catalysis and can be reasonably assigned as a dead-end complex. Much longer simulation runs may reveal other binding poses of fosfomycin that converge quickly toward the productive binding pose, making it nearly impossible to deconvolute the contribution of each pose to catalysis.

Our observations highlight characteristics for IPK that have more general implications, namely the structure of the active site accommodates related reactions with promiscuous substrates, although the promiscuous activity does not always utilize the active site residues in the same manner as the native reaction *(5)*. Moreover, bi- or even multimodal binding of promiscuous substrates is a direct consequence of these characteristics and could give rise to multiple products from a single promiscuous substrate.

Our results suggest a possible scenario for the evolution of a bacterial FomA from an archaeal IPK by horizontal transfer of an archaeal *ipk* gene to *Streptomyces*, followed by adaptive mutations that confer antibiotic resistance. This requires a sufficient level of promiscuous fosfomycin activity of IPK to provide an advantage for selection of FomA activity in the presence of environmental pressure (fosfomycin) *(4)*. IPK is an enzyme found exclusively in the alternate mevalonate route for biosynthesis of IPP in *Archaea*, most of which is incorporated into the phytanyl and diphytanyl chains of their cell membrane phospholipids *(48)*. *Bacteria* and *Eukarya* synthesize IPP by different pathways. Thus, the translocated *ipk* gene initially did not have a metabolic function in the receiving bacterium and was immediately available for specialization as a fosfomycin

resistance gene. Activities with catalytic efficiencies similar to the fosfomycin kinase activity of IPK have been shown to confer immediate selective advantage, as in the case of phosphonate monoester hydrolase in *Burkholderia caryophili* (9.5 $M^{-1}s^{-1}$), which allowed this organism to thrive on glyceryl glyphosate, a previously unencountered phosphorous source *(49-51)*. Even a catalytic efficiency as low as 0.3 $M^{-1}s^{-1}$ (*E. coli* glutamine phosphoribosylpyrophosphate amidotransferase) can provide a similar advantage, although in this organism the gene was overexpressed from a multicopy plasmid under control of a strong promoter *(52)*. Moreover, a phylogenetic tree of IPK homologues clustered the *S. wedmorensis* and *S. fradiae* FomA enzymes with archaeal IPKs, making this case of horizontal gene transfer from *Archaea* to *Bacteria* a viable scenario *(10, Figure 1)*. Likely instances of horizontal transfer of metabolic genes between *Archaea* and *Bacteria* (and other interdomain HGTs) have been pointed out, as in the acquisition of mevalonate pathway genes from *Archaea* (or Eukarya) by *Borrelia* and *Cocci*, or of the bacterial glycerol dehydrogenase gene acquired by *Archaea,* which was possibly co-opted for glycerol-1-phosphate dehydrogenase activity *(53)*.

The substrate promiscuity of IPK towards fosfomycin and the highly conserved catalytic site residues and similar folds of IPK and FomA indicate a genetic connection between the two proteins, even though they only share 20-25% identity at the amino acid level. The fact that IPK is found in all representatives of *Archaea* with sequenced genomes while FomA is only found in two strains of *Streptomyces* and in one strain of *Pseudomonas* suggests that the bacterium acquired an archaeal copy of an IPK gene by horizontal gene transfer, which was subsequently optimized for FomA activity. Our findings underscore the gene duplication/horizontal transfer mechanism for evolution of

new enzyme activities, including new metabolic pathways, from enzymes available in the cell.

References

1. Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signaling. *Nature 411*, 355-365.

2. Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol. 30*, 409-425.

3. Ohno, S. (1970) Evolution by gene duplication. George Allen and Unwin, London.

4. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol. 6*, R91-R105.

5. Copley, S. (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol. 7*, 265-272.

6. Yang, K., and Metcalf, W.W. (2004) A new activity for an old enzyme: Escherichia coli bacterial alkaline phosphatase is a phosphite-dependent hydrogenase. *Proc. Natl. Acad. Sci. 101(21)*, 7919-7924.

7. Miller, B.G., and Raines, R.T. (2005) Reconstitution of a defunct glycolytic pathway via recruitment of ambiguous sugar kinases. *Biochemistry 44*, 10776-10783.

8. Mabanglo, M.F., Schubert, H.L., Chen, M., Hill, C.P., and Poulter, C.D. (2010) X-ray structures of isopentenyl phosphate kinase. *ACS Chemical Biology 5(5)*, 517-527.

9. Pakhomova, S., Bartlett, S. G., Augustus, A., Kuzuyama, T., Newcomer, M. E. (2008) Crystal structure of fosfomycin resistance kinase FomA from Streptomyces wedmorensis, *J. Biol. Chem. 28*3, 28518-28526.

10. Chen, M., and Poulter, C.D. (2010) Characterization of thermophilic archaeal isopentenyl phosphate kinases. *Biochemistry 49*, 207-217.

11. Dela Cruz, F., Garcia-Lobo, J.M., and Davies, J. (2002) Bacterial resistance to antimicrobials. (Lewis, K., Salyers, A. A., Taber, H.W., and Wax, R. G., Eds.) pp 19-36, Marcel Dekker, Inc. New York, NY.

12. Suarez, J.E., and Mendoza, M.C. (1991) Plasmid-encoded fosfomycin resistance. *Antimicrob. Agents Chemother. 35*, 791-795.

13. Bernat, B.A., Laughlin, L.T., and Armstrong, R.N. (1997) Fosfomycin resistance protein (FosA) is a manganese metalloglutathione transferase related to glyoxalase I and the extradiol dioxygenases. *Biochemistry 36*, 3050-3055.

14. Cao, M., Bernat, B.A., Wang, Z., Armstrong, R.N., and Helmann, J.D. (2001) FosB, a cysteine-dependent fosfomycin resistance protein under the control of sigma (W), an extracytoplasmic-function sigma factor in *Bacillus subtilis*. *J. Bacteriol. 183*, 2380-2383.

15. Etienne, J., Gerbaud, G., Courvalin, P., and Fleurette, J. (1989) Plasmid-mediated reistance to fosfomycin in *Staphylococcus epidermidis*. *FEMS Microbiol. Lett. 52*, 133-137.

16. Keller, R. K., and Thompson, R. (1993) Rapid synthesis of isoprenoid diphosphate and their isolation in one step using either thin layer or flash chromatography, *J. Chromatography A 645*, 161-167.

17. Pilloff, D., Dabovic, K., Romanovski, M.J., Nonanno, J.B., Doherty, M., Burley, S.K., and Leyh, T.S. (2003) The kinetic mechanism of phosphomevalonate kinase. *J. Biol. Chem. 278*, 4510-4515.

18. Leatherbarrow, R.J. (2001) GraFit Version 5, Erithacus Software Ltd., Horley, U.K.

19. Zewe, V., and Fromm, H.J. (1965) Kinetic studies of rabbit muscle lactate dehydrogenase. *Biochemistry 4(4)*, 782-792.

20. 20. Lindahl E., Hess B., and van der Spoel D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. (Springer Berlin/Heidelberg), 306-317.

21. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J. Yin, D., and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B, 102*, 3586−3616.

22. V. Zoete, M. A. Cuendet, A. Grosdidier, O. Michielin, *SwissParam, a Fast Force Field Generation Tool For Small Organic Molecules*, to be submitted.

23. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M.

L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926– 935.

24. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J., Kapral, G.J., Wang, X., Murray, L., Arendall, B., III, Snoeyink, J., Richardson, J. S., and Richardson, D.C. (2007) Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res. 35*, W375-W383, Web Server Issue.

25. Essman, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995) A smooth particle mesh Ewald potential. *J. Chem. Phys. 103*, 8577-8592.

26. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys. 126,* 012101.

27. Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. 52*, 7182-7190.

28. Knowles, J.R. (1980) Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem. 49*, 877-819.

29. O'Brian, C.A., and Ward, N.E. (1990) Characterization of calcium- and phospholipid-dependent ATPase reaction catalyzed by rat brain protein kinase C. *Biochemistry 29*, 4278-4282.

30. Pakhomova, S., Bartlett, S., Doerner, P.A., and Newcomer, M. (2011) Structural and biochemical insights into the mechanism of fosfomycin phosphorylation by fosfomycin resistance kinase FomA. *Biochemistry,* just accepted manuscript.

31. Hendlin, D., Stapley, E.O., Jackson, M., Wallick, H., Miller, A. K., Wolf, F.J., Miller, T.W., Chaiet, L., Kahan, F. M., Foltz, E. L., Woodruff, H. B., Mata, J. M., Hernandez, S., and Mochales, S. (1969) Phosphonomycin, a new antibiotic produced by strains of *Streptomyces*. *Science 166*, 122-123.

32. Shoji, J., Kato, T., Hinoo, H., Hattori, T., Hirooka, K., Matsumoto, K., Tanimoto, T., and Kondo, E. (1986) Production of fosfomycin (phosphonomycin) by *Pseudomonas syringae J. Antibiot.* (Tokyo) 39, 1011-1012.

33. Lobel, B. (2003) Short-term therapy for uncomplicated urinary tract infection today. Clinical outcome upholds the theories. *Int. J. Antimicrob. Agents 22* (Suppl. 2), 85-87.

34. Cassone, M., Campanile, F., Pantosti, A., Venditti, M., and Stefani, S. (2004) Identification of a variant "Rome clone" of methicillin-resistant *Staphylococcus aureus* in France: a phylogenetic approach. *Microb. Drug Resist. 10*, 43-49.

35. Nakazawa, H., Kikuchi, Y., Honda T., Isago, T., and Nozaki, M. (2003)

Enhancement of antimicrobial effects of various antibiotics against methicillin-resistant Staphylococcus aureus (MRSA) by combination with fosfomycin. *J. Infect. Chemother. 9*, 304-309.

36. Kahan, F.M., Kahan, J.S., Cassidy, P.J., and Kropp, H. (1974) The mechanism of action of fosfomycin (phosphonomycin) *Ann. N.Y. Acad. Sci. 235,* 364-386.

37. Marquardt, J. L., Brown, E. D., Lane, W. S., Haley, T. M., Ichikawa, Y., Wong, C. H., and Walsh, C. T. (1994) Evidence that the fosfomycin target Cys115 in UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) is essential for product release. *Biochemistry 33*, 10646-10651.

38. Kadner, R. J., and Winkler, H. H. (1973) Isolation and characterization of mutations affecting the transport of hexose phosphates in Escherichia coli. *J Bacteriol. 113*, 895-900.

39. Tsuruoka, T., and Yamada, Y. (1975) Characterisation of spontaneous fosfomycin (phosphonomycin)-resistant cells of *Escherichia coli* B "in vitro". *J. Antibiot.* (Tokyo) *28*, 906-911.

40. Nair, S. K., and van der Donk, W. A. (2010) Structure and mechanism of enzymes involved in biosynthesis and breakdown of the phosphonates fosfomycin, dehydrophos, and phosphinothricin. *Arch. Biochem. Biophys. 505(1)*, 13-21.

41. Kobayashi, S., Kuzuyama, T., and Seto H. (2000) Characterization of the *fomA* and *fomB* gene products from *Streptomyces wedmorensis*, which confer fosfomycin resistance on *Escherichia coli*. *Antimicrob. Agents Chemother. 44*, 647-650.

42. Woodyer, R. D., Shao, Z. Thomas, P. M., Kelleher, N. L., Blodgett, J. A., Metcalf, W. W., van der Donk, W. A., and Zhao, H. (2006) Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. *Chem. Biol. 13*, 1171-1182.

43. Marco-Marin, C., Gil-Ortiz, F., and Rubio, V. (2005) The crystal structure of *Pyrococcus furiosis* UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis, *J. Mol. Biol. 352*, 438-454.

44. Hall, A., and Knowles, J.R. (1976) Directed selective pressure on a -lactamase to analyse molecular changes involved in development of enzyme function. *Nature 264*, 803-804.

45. Healey, W. J., Labgold, M. R., and Richards, J. H. (1989) Substrate specificities in class A -lactamases: Preference for penams vs. cephams. The roles of residue 237. *Proteins 6*, 275-283.

46. Gibson, R. M., Christensen, H., and Waley, S. G. (1990) Site-directed mutagenesis of -lactamase I. Single and double mutants of Glu-166 and Lys73. *Biochem. J. 272*, 613-619.

47. Imtiaz, U., Manavathu, E. K., Lerner, S. A., and Mobashery, S. (1993) Critical hydrogen bonding by Ser235 for cephalosporinase activity of TEM-1 -lactamase. *Antimicrobial Agents and Chemotherapy 37*, 2438-2442.

48. Boucher, Y. (2007) Lipids: biosynthesis, function, and evolution in *Archaea*, Molecular and Cellular Biology (Cavicchioli, R., Ed.) pp 341-353 ASM Press, Washington, DC.

49. Van Loo, B., Jonas, S., Babtie, A. C., Benjdia, A., Berteau, O., Hyvönen, M., and Hollfelder, F. (2010) An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily, *Proc. Natl. Acad. Sci.* ***2010***, *107 (7),* 2740-2745.

50. Ghanem, E., Li, Y., Xu, C., and Raushel, F. M. (2007) Characterization of a phosphodiesterase capable of hydrolyzing EA 2192, the most toxic degradation product of the nerve agent VX. *Biochemistry 46*, 9032-9040.

51. McLoughlin, S. Y., Jackson, C., Liu, J. W., and Ollis, D. L. (2004) Growth of *Escherichia coli* co-expressing phosphotriesterase and glycerophosphodiester phosphodiesterase, using paraoxon as the sole phosphorous source. *Appl. Environ. Microbiol. 70*, 404-412.

52. Patrick, W. M., and Matsumura, I. (2008) A study in molecular contingency: glutamine phosphoribosylphosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol. 377*, 323-336.

53. Boucher, Y., Kamekura, M., and Doolittle, W. F. (2004) Origins and evolution of isoprenoid lipid biosynthesis in Archaea. *Mol. Microbiol. 52(2)*, 515-527.

CHAPTER 4

ISOPENTENYL PHOSPHATE KINASE MUTANTS WITH GERANYL PHOSPHATE AND FARNESYL PHOSPHATE KINASE ACTIVITIES

Introduction

Directed evolution and structure-based redesign are two approaches whereby wild-type enzymes are conferred with novel activities *(1)*. Well-documented examples of these protein engineering strategies are found in the literature, resulting in mutants with improved catalytic efficiencies and selectivities for nonnatural substrates. In the case of prenyltransferases and isoprenoid synthases, several studies have been performed in which their chain length selectivities and ultimate product chain lengths were altered. For instance, Ohnuma *et al.* used a random chemical mutagenesis approach to generate mutants of *Bacillus stearothermophilus* farnesyl diphosphate synthase (FPPS) with geranylgeranyl diphosphate synthase (GGPPS) activity *(2)*. The selection was accomplished by monitoring the formation of GGPP *in vivo* using the *Erwinia uredovora* carotenoid biosynthesis enzymes, phytoene synthase and phytoene desaturase, such that successful mutant colonies displayed a red coloration coming from lycopene. The study resulted in the identification of an aromatic active site residue, Tyr81, whose mutation to a smaller side chain (histidine) was essential for GGPP synthase activity. This change in

ultimate product chain length specificity came at the expense of catalytic activity, so that mutants harboring the Tyr81His mutation have GPPS activities lower than the wild-type FPPS activity. Interestingly, it was also found that the Michaelis constant, K_M, of IPP changed depending on the allylic primers used, indicating that the mutant enzymes possessed different affinities to the same substrate depending on the countersubstrate, a phenomenon that might arise from conformational changes upon binding of the latter *(2)*.

Follow-up experiments by Ohnuma *et al.* corroborated the importance of the position occupied by Tyr81 in *B. stearothermophilus* FPPS in determining ultimate chain length specificity *(3)*. Comparison of isoprenoid synthase enzymes from different organisms showed that the corresponding residues occupy the fifth position upstream of the first DD*XX*D motif, one of two aspartate-rich sequences that serve as binding sites for the diphosphate moieties of IPP and allylic substrates *(4-16)*. The fifth position is occupied by aromatic side chains (phenylalanine and tyrosine) in archaeal FPPS and GGPPS. On the other hand, polyprenyl diphosphate synthases that catalyze the condensation beyond GGPP in organisms other than *Archaea* have a non-aromatic amino acid at this position. Thus, transmutation of Phe77, which occupies the fifth position in *Sulfolobus acidocaldarius* GGPPS, to either serine or leucine resulted in mutants that can catalyze the condensation of IPP to form geranylfarnesyl diphosphate (GFPP) and hexaprenyl diphosphate (HPPS) *(2)*. These mutants were generated and identified using random chemical mutagenesis and a complement system of a yeast mutant deficient in HPPS activity. Further studies by *Ohnuma et al.* on the S. *acidocaldarius* GGPPS showed the fifth position before the first DD*XX*D motif to be but a part of a larger region that determines ultimate chain length specificity, thereby called the chain length

determination (CLD) region that starts at the fifth position up to the end of the first asparate-rich motif *(17)*. This CLD region was previously shown by Tarshis *et al.* to form an α helix, where the distance between the aspartate (Asp177) in the first DD*XX*D motif and the amino acid in the fifth position is about 10-12 Å, similar to the length of the FPP hydrocarbon chain *(18)*. Indeed, the CLD region blocks further elongation beyond the limit of the chain length of product determined by the specificity of each enzyme.

The three-dimensional structure of the avian FPPS also guided the construction of mutants with GPPS activity. Stanley Fernandez *et al.* used modeling and mutagenesis studies to identify the Ala116Trp mutation adjacent to Asp117 of the first DD*XX*D motif that caused the highest improvement in GPPS activity among seven designed mutants *(19)*. Moreover, a similar level of activity was found for the Asn144'Trp mutant, although in this instance Asn144 is found across the binding pocket from Ala116 on the other polypeptide chain in the homodimer. These two mutants possessed 63- and 28-fold improvements in k_{cat}/K_M for the synthesis of GPP relative to FPP. These examples thus demonstrate the utility of directed evolution and structure-based engineering in expanding the activities of isoprene synthases.

In another case, structure-based engineering was exploited in the conversion of the chain length specificity of prenyltransferases. These enzymes catalyze the attachment of lipid moieties to proteins. This process, called prenylation, is important in the localization of target proteins to the cellular membrane. Eukaryotic cells harbor three kinds of protein prenyltransferases: protein farnesyltransferase (pFTase), protein geranylgeranyl transferase type I (pGGTase I) and protein geranylgeranyl transferase type II (pGGTase II or Rab GGTase), whose substrates are limited to the Rab subfamily

of G proteins *(20)*. On the other hand, pFTase and pGGTase I prenylate proteins that include members of the Ras superfamily of GTPases, several protein kinases and phosphatases, and several proteins involved in centromere function and nuclear integrity *(21,22)*. pFTase in particular transfers the 15-carbon farnesyl chain while pGGTase I transfers the 20-carbon geranygeranyl chain to the cysteine residue of a C-terminal Ca_1a_2X motif of the substrate protein *(20)*. In this motif, C refers to the cysteine residue, a_1 and a_2 represent any small aliphatic residue, and X determines the whether the protein is farnesylated or geranylgeranylated *(23)*. After prenylation, the a_1a_2X sequence is excised by an endoprotease, and the prenylcysteine residue is carboxymethylated by Icmt *(24)*. The resulting prenylated protein has high affinity for cellular membranes and possesses a unique structure that can serve as a specific recognition factor in protein-protein interactions *(25-27)*.

Prenyltransferases are highly selective for their cognate isoprenoid diphosphates, and crystal structures have shown that these substrates bind along one side of the active site, with the diphosphate bound in a positively charged cleft in the pocket and the lipid tail extending deep into the pocket *(21,28-37)*. Crystal structures of pGGTase I with GGPP and GGPP analogues reveal that the first three isoprene units bind in the pocket in the same way that the lipid moiety of FPP binds in the pFTase active site, while the fourth isoprene unit sits at a 90° angle relative to the rest of the lipid molecule, across the bottom of the active site *(21,36,38)*. Moreover, the binding site for the first three isoprene units are essentially conserved in pFTase and pGGTase I. Placement of the observed binding conformation of GGPP in pGGTase I to the corresponding site in pFTase results in a steric clash with Trp102 and Tyr365, suggesting that these two residues are the key

determinants of isoprenoid diphosphate specificity. Terry *et al.* subsequently showed that a Trp102Thr mutant, given the same threonine residue found in the corresponding position in pGGTase I, completely switched the specificity of pFTase from FPP to GGPP *(39)*. Moreover, the turnover rate for the geranylgeranyl transfer reaction catalyzed by the Trp102Thr mutant is similar to that observed for farnesyl transfer by native pFTase. The mutation essentially abolished the use of FPP but retained the specificity for pFTase Ca_1a_2X protein substrates. Thus, through a single mutation, the isoprenoid specificity of pFTase was shifted without sacrificing catalytic efficiency. On the other hand, mutation of Tyr365 in pFTase to the corresponding residue in pGGTase I (phenylalanine) relaxed the isoprenoid selectivity of the mutant in that it could use both FPP and GGPP to prenylate typical pFTase protein substrates. Finally, a combination of the two pFTase mutations Trp102Thr and Tyr365Phe resulted in a variant with a kinetic profile similar to the one harboring only the Trp012Thr mutation. This mutant, as well as the Trp102Thr pFTase variant, could be used to elucidate the *in vivo* effects of alternate prenylation on cellular localization and function *(39)*. Thus, structure-guided engineering of enzymes presents a powerful way to create proteins with new activities intended for practical uses.

The structures of IPK discussed in Chapter 2 revealed the amino acid residues that form the hydrophobic binding pocket of IP *(40)*. Dellas and Noel first demonstrated in MJ IPK that transmutation of these side chains into less bulky ones, either singly or in combination, resulted in enzymes capable of phosphorylating the longer chain isoprenoid farnesyl phosphate *(41)*. They also observed that the mutations are contextually dependent in that mutations in the back of the IP binding site were not effective without mutations at the front. The MJ IPK mutants were not kinetically characterized and their

catalytic efficiencies for phosphorylation of farnesyl phosphate and geranyl phosphate are not known.

In vitro enzymatic studies by Thai *et al.* demonstrated the presence of kinase activities from *Nicotiana tabacum* microsomal fractions capable of converting farnesol to farnesyl diphosphate through successive monophosphorylation reactions *(42)*. The evidence for an FP kinase is based on demonstrating that exogenous FP is converted to FPP by a CTP-dependent reaction, corroborated by the observation that CTP could be reversibly formed when microsomes were incubated with CDP and FPP or GGPP. Other farnesol kinases capable of producing FPP through a farnesyl monophosphate intermediate have been observed in 100,000 × g pellets from microalga *(43)*, cell extracts from a halophilic archaeon *(44)*, and microsomes from rat liver *(45,46)*. In this chapter, several mutants of THA IPK with FP kinase activities, in conjunction with more pronounced geranyl phosphate (GP) kinase activities, are presented. Product analyses and steady-state kinetic characterizations were performed on these mutants, constituting the first direct demonstration of GP kinase activity to date. The work contained in this last chapter thus serves as a prelude to the creation of isoprenyl monophosphate kinases capable of labeling isoprenoid substrates with ^{32}P, with applications in research involving isoprenoids and their derivatives.

Experimental Procedures

Materials. γ-[^{32}P] ATP was purchased from American Radiolabeled Chemicals. Tricholoroacetonitrile (>95%), farnesol (95%), geraniol (98%), phosphoenolpyruvate (PEP) and nicotinamide adenine dinucleotide (NADH) were purchased from Sigma-

Aldrich. Phosphoric acid (85%) was purchased from Reagents. Bovine serum albumin was purchased from Invitrogen. LDH (rabbit muscle) and PK (rabbit muscle) were purchased from Roche. The concentrations of all reagents containing phosphates were determined by phosphate analysis, and the precise concentration of mutant IPK enzymes were measured using the BCA Protein Assay Kit from Pierce.

Synthesis of GP (2) and FP (3). The procedure by Keller *et al.* was used to synthesize GP and FP *(47)*. Briefly, a dry solution of 3.46 g (17.4 mmol) triethylammonium phosphate (TEAP) in 15 mL of acetonitrile was added dropwise to a stirred mixture of alcohol (geraniol or farnesol, 3.25 mmol) in 5 mL of trichloroacetonitrile at room temperature under N_2, in three equal portions with a 5 min interval between additions. The mixture was then concentrated at reduced pressure, giving a thick, dark orange oil. The resulting oil was chromatographed on silica gel using 2-propanol/NH_4OH/H_2O (6:3:1 v/v/v) and fractions containing pure products were combined. The solvent was removed by lyophilization, resulting in a white solid with a yield of 18%. Detection of GP and FP was performed by thin layer chromatography using the same mobile phase and stained with *p*-anisaldehyde, R_f (GP) = 0.47, R_f (FP) = 0.5. ^{1}H NMR (D_2O) for GP: δ 1.59 (s, 3H), 1.65 (s, 3H), 1.68 (s, 3H) 2.01-2.18 (m, 4H), 4.38 (m, 2H), 5.17 (t, 1H, *J*=13.5 Hz), 5.42 (t, 1H, *J*=7.5 Hz); ^{31}P NMR (D_2O) for GP: δ 2.20. ^{1}H NMR (D_2O) for FP: δ 1.47 (s, 3H), 1.49 (s, 3H) 1.54 (s, 3H), 1.61 (s, 3H), 1.82-2.06 (m, 8H), 4.35 (m, 2H), 4.95-5.07 (m, 2H), 5.34 (t, J=6.7 Hz); ^{31}P NMR (D_2O) for FP: δ 3.94. The masses of the products were confirmed by major peaks for $[GP\text{-}H^+]^-$ (233.1) and $[FP\text{-}Na^+]^-$ (325.2) or $[FP\text{-}H^+]^-$ (301.1).

Mutation of THA IPK. To create THA IPK mutants, computer modeling was performed using the UCSF CHIMERA software to dock GP and FP in the active site of THA IPK with *in silico* mutations *(48)*. Mutations to create the desired mutants were then carried out using the QuikChange Lightning Multi-Site Directed Mutagenesis Kit (Agilent Technologies) and mutagenic primers designed using the Quikchange Primer Design Tool. The plasmid template contained the THA IPK gene in the pET28b vector to include the N-terminal sequence MGSSHHHHHHSSGLVPRGS upstream of the IPK sequence. Correct mutant sequences were confirmed by sequencing at the University of Utah DNA Sequencing Core Facility.

Expression and purification of THA IPK mutants. Plasmids of mutant THA IPKs were chemically transformed into BL21 (DE3) Star cells (Invitrogen) and plated on agar containing kanamycin. Mutant colonies were used to make 10 mL starter cultures in LB broth, which were then inoculated into 1 L of LB medium. The cultures were shaken at 37 °C until OD_{600} reached 0.5 and induced with 1 mM of IPTG followed by incubation overnight at 30 °C. Cells were harvested by centrifugation and stored at -80 °C until purification. For purification of THA IPK mutants, see Experimental Procedures in Chapter 3.

Detection of GP and FP kinase activities using γ-[^{32}P] ATP. Purified THA IPK mutants were tested for GP, FP and IP kinase activities by incubating 10 μM of each enzyme with 5 mM isoprenoid substrate in 100 mM HEPES buffer, pH 7.5, containing 10 mM β-mercaptoethanol, 10 mM $MgCl_2$, 1 mg/mL BSA, 1 μM γ-[^{32}P] ATP for 2 hr at 37 °C in a total volume of 50 μL. The mixtures were quenched with 100 μL of 500 mM EDTA. A 10 μL to 20 μL portion of each incubate was spotted on silica plates and

developed with $CHCl_3$/pyridine/formic acid/ H_2O (30:70:16:10 v/v/v/v). The TLC plate was imaged for 24 h using a storage phosphor autoradiography cassette and visualized using a Typhoon 8600 variable mode imager (GE Healthcare).

Kinetic characterization of THA IPK mutants. The protocol for fluorescent assays was based on the procedure by Pilloff *et al.* with slight modifications *(49)*. The activities of the coupling enzymes, PK and LDH, were measured using the procedure in Chapter 3. To initiate the reaction, a THA IPK mutant was added to the assay buffer (100 mM HEPES, pH 7.5, containing 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mg/mL BSA and 250 μM ATP (saturating)), including appropriate amounts of GP, FP or IP and coupling enzymes in a final volume of 200 μL. The reaction was monitored at 37 °C for 600 s by observing the change in fluorescence (λ_{ex} = 340 nM, λ_{em} = 460 nm) (FluroMax, Jobin Yvon Horiba). The initial rates were measured from the linear portion of the curve (< 15% consumption of the concentration limiting substrate). The kinetic constants were determined by fitting initial rates to equation 1 using nonlinear regression in GraFit5 *(50, Erithacus Software)*

$$V/[E] = k_{cat}\,[S] / [S] + K_M \qquad (1)$$

where V is the initial rate, [E] is the total concentration of enzyme in the mixture, [S] is the concentration of the isoprenoid substrate, and K_M is the Michaelis constant. The concentration of ATP was chosen after performing similar experiments using saturating concentrations of the isoprenoid substrate.

UPLC-MS of THA IPK mutant products, GPP and FPP. Samples for UPLC-MS of GPP, FPP and IPP produced by THA IPK mutants were obtained by incubating 10 uM of each mutant with 2 mM of GP, FP, and IP (as control) in 100 mM HEPES buffer, pH 7.5, containing 100 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mg/mL BSA and 5 mM ATP at 37 °C for 2 h. The mixtures were then centrifuged at 4 °C and 3000 rpm to remove the enzyme using a 10,000 MWCO Centricon (Millipore). The collected fractions were flash frozen and lyophilized overnight, then dissolved in minimum volume of 25 mM NH_4HCO_3. Isoprenoid diphosphate products (GPP, FPP and IPP) were separated from substrates (GP, FP and IP) on a C18 column (for GPP and FPP) and a C4 column (for IPP) on an Acquity UPLC. Isocratic elution with 95% 25 mM NH_4HCO_3 and 5% acetonitrile was used to separate GPP and GP, while isocratic elution with 90% 25 mM NH_4HCO_3 and 10% acetonitrile was used to separate FPP and FP, each at a flow rate of 0.6 mL/min. For the separation of IPP and IP, isocratic elution with 100% of 25 mM NH_4HCO_3 was performed. Peaks with masses corresponding to each substrate and product were detected by negative-ion (ES-) MS.

Results

Mutation of THA IPK and qualitative GP and FP kinase activity assays. A total of 8 mutants were created, initially using the pET151/D-TOPO THA IPK template from Chapter 2 used to crystallize and determine the structure of THA IPK. GP and FP in fully extended conformations were used in the *in silico* modeling for the identification of candidate mutations. The THA IPK variants contained different combinations of alanine mutations of the following residues in the periphery of the IP binding site: Leu58, Pro59,

Tyr70, Val73, Val130 and Ile140 *(40)*. The first two residues are found on the β3-β4 hairpin that caps the IP binding site, Tyr70 and Val73 are found on the long αC helix, and Val130 and Ile140 are on the β9 and β10 strands, respectively (Figure 4.1a and b). Among these residues, Val73, Val130 and Ile140 are found at the front of the IP binding site, and based on inspection of the structure, mutation of these residues appeared to be necessary in order for mutations in the back to have any effect on the presumed binding of GP and FP in the enlarged pocket. The Tyr70Ala mutation appeared to be important to accommodate a fully extended FP molecule, while the Leu58Ala and Pro59Ala mutations were presumed to allow the binding of FP in the absence of the Tyr70Ala mutation. Table 4.1 lists the different mutants initially tested for GP and FP kinase activities. In the initial GP and FP kinase activity assays, crude purifications of these mutants were performed, owing to the difficulty of the pET151/D-TOPO constructs to yield homogeneous proteins. Moreover, the mutants appeared to express in different yields, such that in the initial activity assays, the total concentrations of mutant proteins used were not the same. Thus, the GP and FP kinase assays were qualitative estimations of activity (Figure 4.2a-4f). Based on the initial results of the GP and FP kinase activity assays, four mutants were chosen, namely Mutant 12, Mutant X, Mutant 7 and Mutant 11. These mutants appeared to have the strongest GP kinase activities among the 8 mutants tested. Mutant 12, Mutant X, and Mutant 7 contain the Tyr70Ala mutation and were presumed to bind a fully extended FP molecule. Mutant 11 does not contain the Tyr70Ala mutation and was expected to have minimal FP kinase activity through binding FP in a nonlinear conformation. Among the four mutants, Mutant 12 has the biggest

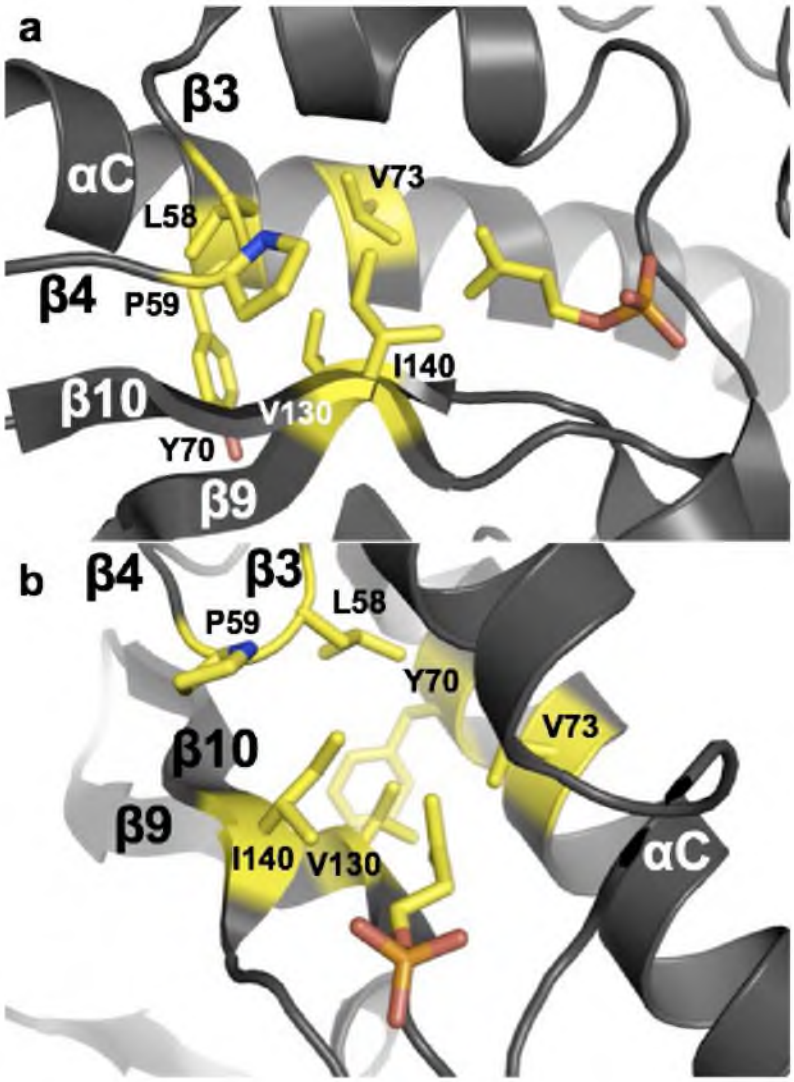


Figure 4.1. The THA IPK active site showing residues in the IP binding site mutated to alanine. (a) Side view of the THA IPK N-terminal lobe showing secondary structure elements that form the walls of the IP binding site and the chain length limiting residues. (b) Alternate view of the IP binding site.

Table 4.1. List of engineered THA IPK mutants.

Mutant name	Combination of residues mutated to alanine
Mutant 6	Tyr70, Val130
Mutant 7	Tyr70, Val130, Ile140
Mutant 8	Leu58, Pro59, Tyr70, Val130, Ile140
Mutant 9	Leu58, Pro59, Tyr70, Val73, Val130, Ile140
Mutant 10	Leu58, Pro59, Val73, Val130, Ile140
Mutant 11	Val73, Val130, Ile140
Mutant 12	Tyr70, Val73, Val130 Ile140
Mutant X	Tyr70, Val73, Ile140

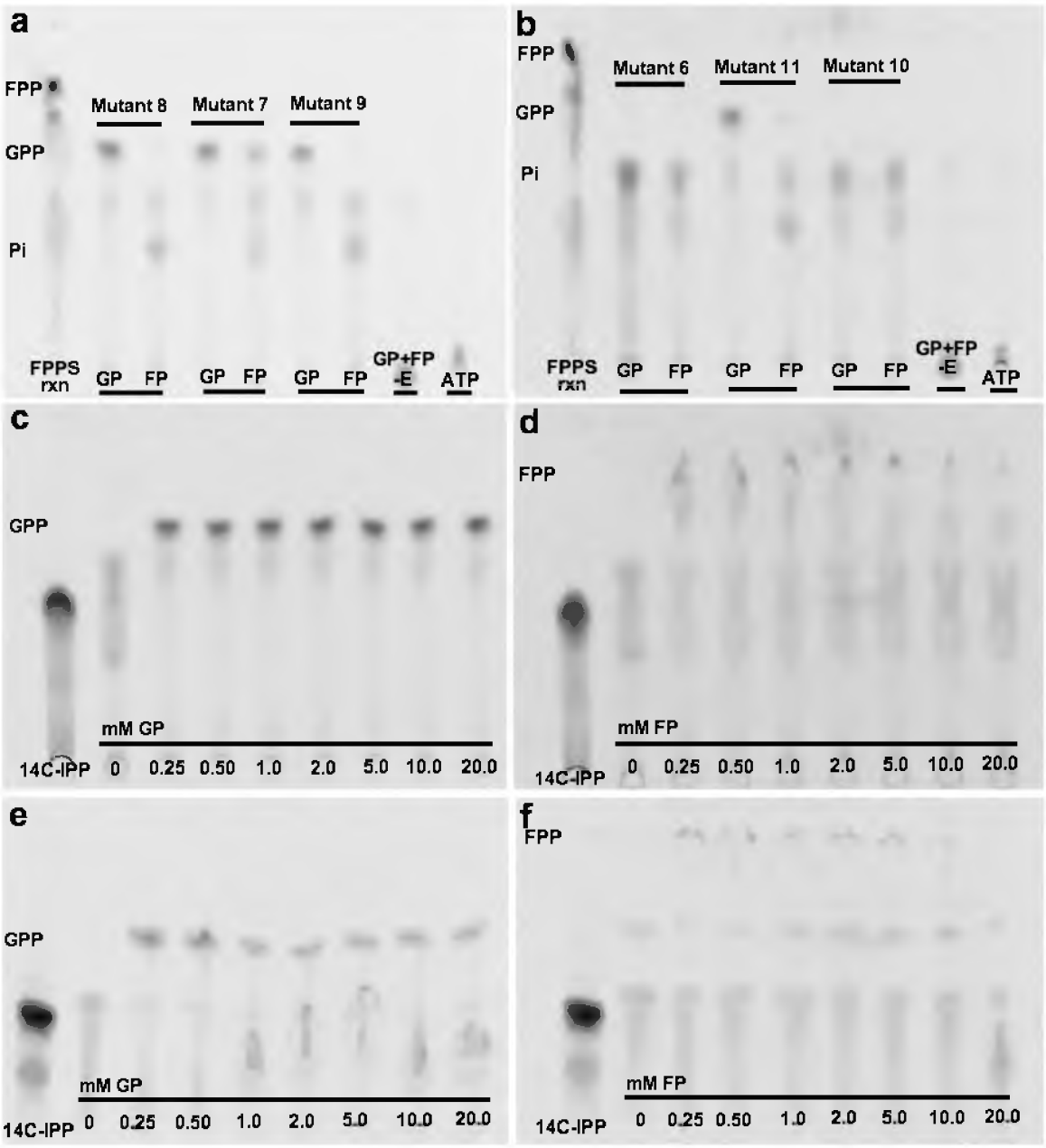


Figure 4.2. Autoradiogram showing product formation by THA IPK mutants incubated with GP, FP, and γ-[^{32}P] ATP. (a-b) Mutants 8, 7, 9, 6, 11 and 9 incubated with 10 mM GP and FP. The radioactive markers are the products produced by the incubation of avian FPP synthase with ^{14}C-IPP and DMAPP. (c-d) Mutant X incubated with increasing concentrations of GP (c) and FP (d) for 2 hr at 37 °C. (e-f) Mutant 12 incubated with increasing concentrations of GP (e) and FP (f) for 2 hr at 37 °C.

substrate binding pocket as it contained all four Tyr70Ala, Val73Ala, Val130Ala and Ile140Ala mutations found in combinations of three mutations in the other three variants.

In order to obtain homogeneous enzymes for kinetic characterization, Mutant 12, Mutant X, Mutant 7, and Mutant 11 were subcloned into the pET28b vector, expressed and purified (Figure 4.3a). GP and FP kinase activity for each mutant were again tested (Figure 4.3b). All four mutants exhibited strong GP kinase activities, as shown by a strong radioactive spot corresponding to GPP (R_f = 0.52). In addition, Mutant 12, Mutant X, and Mutant 7, each containing the Tyr70Ala mutation, showed detectable FP kinase activities (lighter radioactive spot, R_f= 0.75) while Mutant 11 (no Tyr70Ala mutation) did not. For all mutants, GP kinase activity appeared stronger than either FP kinase activity or the residual IP kinase activity.

Kinetic characterization of mutants and spectroscopic determination of their GPP and FPP products. The coupled fluorescence assay by Pilloff *et al.* was used to measure the GP and FP kinase activities of Mutant 12, Mutant X, Mutant 7 and Mutant 11 *(49).* Table 4.2 summarizes the steady-state kinetic constants of all four mutants. The kinase reaction is bisubstrate and k_{cat} and K_M for one substrate were measured at a saturating concentration of cosubstrate.

Mutant 12 (Tyr70Ala, Val73Ala, Val130Ala, Ile140Ala). Mutant 12 exhibited improved GP kinase activity compared to wild-type IPK. Figure 4.4a shows a GP molecule docked in the active site of a modeled Mutant 12. For this mutant, k_{cat} = 1.1 (± 0.6) s^{-1} represents a 22-fold increase relative to the GP kinase activity of wild-type THA IPK. In addition, K_M^{GP} = 2.4 (± 0.3) × 10^3 μM for Mutant 12, a slight change from K_M^{GP} for the wild-type enzyme. The catalytic efficiency k_{cat}/K_M = 4.7 × 10^2, for the GP kinase

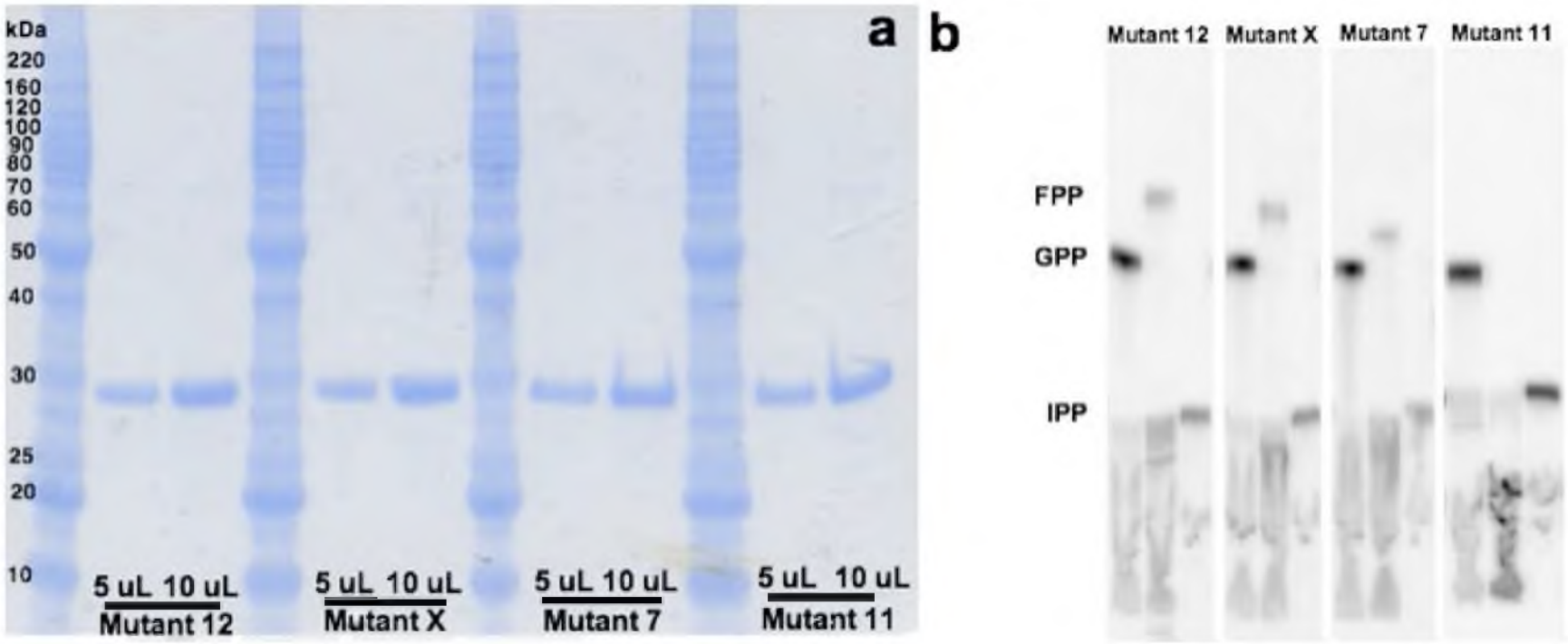


Figure 4.3. Purification of Mutant 12, Mutant X, Mutant 7 and Mutant 11 and product assay using γ-[^{32}P] ATP. (a) Denaturing SDS PAGE showing the four THA IPK mutants purified to homogeneity prior to use in enzyme kinetics. The masses of molecular weight markers are indicated. Monomeric THA IPK mutants have masses ~ 29 kDa. (b) Autoradiogram showing production of GPP and FPP by mutants, as well as the residual IP kinase activity forming IPP. The mutants were each incubated with 5 mM of isoprenoid monophosphate.

Table 4.2. Kinetic constants for the kinase activities of THA IPK mutants.

	WT THA IPK	Mutant 12	Mutant X	Mutant 7	Mutant 11
k_{cat}^{GP} (s^{-1})	0.05	1.11 (± 0.06)	4.04 (± 0.16)	10.08 (± 0.71)	2.80 (± 0.45)
k_{cat}^{FP} (s^{-1})	N. D.	0.62 (± 0.01)	1.40 (± 0.07)	1.48 (± 0.05)	N. D.
k_{cat}^{IP} (s^{-1})	8.0	2.60 (± 0.10) × 10^{-3}	1.01 (± 0.60) × 10^{-2}	4.80 (± 0.30) × 10^{-3}	1.09 (± 0.05) × 10^{-2}
K_M^{GP} (μM)	4.7 (± 1.3) × 10^3	2.35 (± 0.34) × 10^3	3.50 (± 0.34) × 10^3	7.96 (± 1.1) × 10^3	9.50 (± 2.67) × 10^3
K_M^{FP} (μM)	N. D.	1.82 (± 0.13) × 10^3	1.56 (± 0.26) × 10^3	1.50 (± 0.18) × 10^3	N. D.
K_M^{IP} (μM)	4.4 (± 0.5)	5.25 (± 0.75) × 10^3	7.90 (± 0.12) × 10^3	5.66 (± 0.88) × 10^3	4.24 (± 0.68) × 10^3
k_{cat}/K_M^{GP} ($M^{-1}s^{-1}$)	10.0	4.72 × 10^2	1.2 × 10^3	1.3 × 10^3	2.96 × 10^2
k_{cat}/K_M^{FP} ($M^{-1}s^{-1}$)	N. D.	3.40 × 10^2	8.83 × 10^2	9.87 × 10^2	N. D.
k_{cat}/K_M^{IP} ($M^{-1}s^{-1}$)	1.8 × 10^6	0.50	1.28	0.85	2.57
k_{cat}^{ATP} (GP) (s^{-1})	N. D.	1.45 (± 0.06)	1.45 (± 0.04)	1.18 (± 0.03)	1.32 (±0.03)
k_{cat}^{ATP} (FP) (s^{-1})	N. D.	1.26 (± 0.04)	1.65 (± 0.06)	1.10 (± 0.01)	N.D.
k_{cat}^{ATP} (IP) (s^{-1})	6.3 (± 2.2)	1.09 (± 0.02)	1.44 (± 0.05)	1.06 (± 0.02)	1.48 (± 0.07)
K_M^{ATP} (GP) (μM)	N. D.	1.54 (± 0.45)	0.93 (± 0.23)	1.14 (± 0.23)	1.24 (± 0.22)
K_M^{ATP} (FP) (μM)	N. D.	1.32 (± 0.30)	1.36 (± 0.37)	1.16 (± 0.60)	N. D.
K_M^{ATP} (IP) (μM)	6.0 (± 0.5)	0.98 (± 0.18)	1.28 (± 0.29)	1.01 (± 0.12)	1.89 (± 0.49)
k_{cat}/K_M^{ATP} (GP) ($M^{-1}s^{-1}$)	N. D.	9.44 × 10^5	1.57 × 10^6	1.03 ×10^6	1.06 × 10^6
k_{cat}/K_M^{ATP} (FP) ($M^{-1}s^{-1}$)	N. D.	9.51 × 10^5	1.21 × 10^6	9.44 × 10^5	N. D.
k_{cat}/K_M^{ATP} (IP) ($M^{-1}s^{-1}$)	1.3 × 10^6	1.34 × 10^6	1.13 × 10^6	1.05 × 10^6	7.9 × 10^5

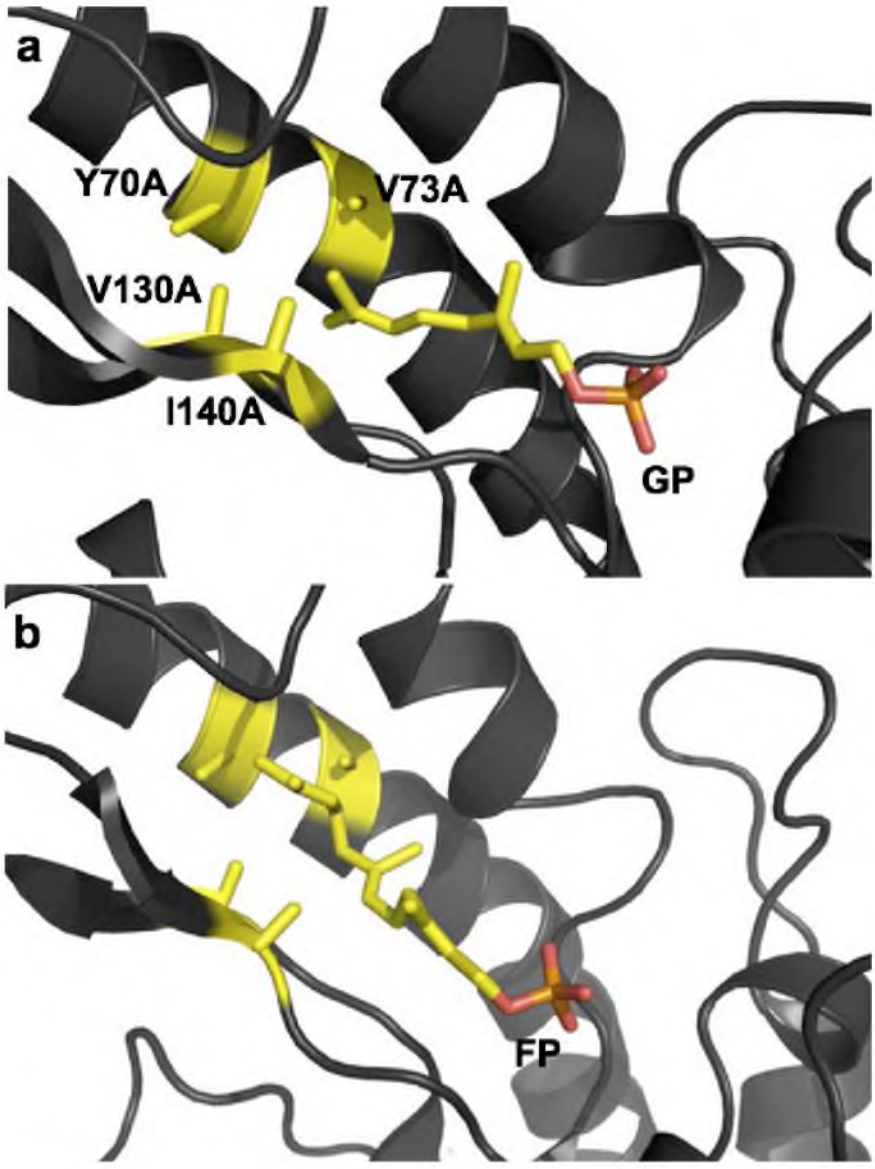


Figure 4.4. Modeled GP (a) and FP (b) molecule in the Mutant 12 active site.

activity of Mutant 12, almost 50-fold larger than the GP kinase activity of the wild-type enzyme.

Figure 4.4b shows an FP molecule modeled in the active site of Mutant 12. The $k_{cat}^{FP} = 6.2 \times 10^{-1}$ s^{-1} for phosphorylation by this mutant is two times lower than k_{cat}^{GP} by the same enzyme. Since the GP kinase activity of wild-type THA IPK was very low, it was presumed that the promiscuous FP kinase activity of wild-type THA IPK would be negligible in comparison. The K_M^{FP} = 1.8 (± 0.1) × 10^3 μM for Mutant 12 to give a catalytic efficiency of 3.4×10^2 $M^{-1}s^{-1}$, which is comparable to the GP kinase activity of Mutant 12. This mutant also exhibited a measurable yet weak residual IP kinase activity, $k_{cat}^{IP} = 2.6 \times 10^{-3}$ s^{-1}, which is 3100-fold lower than k_{cat}^{IP} for wild-type IPK. Moreover, K_M^{IP} = 5.2 (± 0.8) × 10^3 μM for Mutant 12 is 1200-fold higher than K_M^{IP} for wild-type THA IPK, resulting in $k_{cat}/K_M = 5.0 \times 10^{-1}$ $M^{-1}s^{-1}$, which is 3.6×10^6 - fold lower than for wild-type THA IPK. The Michaelis-Menten curves for GP, FP and IP are found in the Appendix.

The k_{cat}^{ATP} and K_M^{ATP} for Mutant 12 using GP, FP, and IP as cosubstrates are similar to those for wild-type THA IPK (Table 4.2), leading to catalytic efficiencies k_{cat}^{ATP}/K_M^{ATP} similar to those of the wild-type enzyme. The Michaelis-Menten curves for ATP in the presence of GP, FP, and IP as cosubstrates are found in the Appendix.

The ability of Mutant 12 to produce the isoprenoid diphosphate products from their respective isoprenoid monophosphate substrates was confirmed by ultra perfomance liquid chromatography and negative ion mass spectrometry. GPP produced by Mutant 12 is evidenced by a sharp peak with a retention time of 1.11 min on a C18 column (See Appendix, M12 GPP). The GPP product was separated from GP substrate, shown by a

smaller peak with a retention time of 3.39 min. The ratio of the peaks supports the high GP kinase activity of Mutant 12. The peaks correspond to a mass of 313 for the negative ion form of GPP ($C_{10}H_{19}O_7P_2^-$) and 233 for the negative ion form of GP ($C_{10}H_{18}O_4P^-$).

Production of FPP by Mutant 12 is evidenced by a smaller peak with a retention time of 4.74 min, supported by a mass of 381 for its negative ion form ($C_{15}H_{27}O_7P_2^-$). The substrate FP has a retention time of 5.43 min, and is supported by a mass of 301 for its negative ion form ($C_{15}H_{26}O_4P^-$) (see Appendix). Finally, we used a C4 column to separate IPP from IP in the incubation mixture with no success. Both compounds were found in the void volume with retention time of less than 1 min, but masses for the negative ion forms of IPP (245, $C_5H_{11}O_7P_2^-$) and IP (165, $C_5H_{10}O_3P^-$) were detected (see Appendix).

Mutant X (Tyr70Ala, Val73Ala, Ile140Ala). Mutant X exhibited improved GP kinase activity compared to Mutant 12 with a k_{cat}^{GP} = 4.1 (± 0.2) s^{-1}, which is 90-fold greater than k_{cat}^{GP} for wild-type THA IPK, and half of k_{cat}^{IP} for wild-type THA IPK. The Michaelis constant, K_M^{GP} = 3.5 (± 0.3) × 10^3 μM, is similar to that for Mutant 12 and wild-type THA IPK, resulting in a catalytic efficiency, k_{cat}/K_M = 1.2 × 10^3 $M^{-1}s^{-1}$ for GP kinase activity that is 3-fold and 120-fold greater than those of Mutant 12 and wild-type THA IPK, respectively. This GP kinase efficiency is only 10^3 lower than that of the native IP kinase activity of THA IPK and represents a significant activity for GP relative to native enzymes. The FP kinase activity of Mutant X is similar to that of Mutant 12, where k_{cat}^{FP} = 1.4 (± 0.7) s^{-1} and K_M^{FP} is 1.6 (± 0.3) × 10^3 μM, leading to k_{cat}/K_M = 8.8 × 10^2 $M^{-1}s^{-1}$ only slightly lower than that of Mutant 12. Figures 4.5a and 4.5b show GP and FP molecules docked in the binding site of Mutant X. The Michaelis-Menten curves for the GP and FP kinase activities of Mutant X are found in the Appendix.

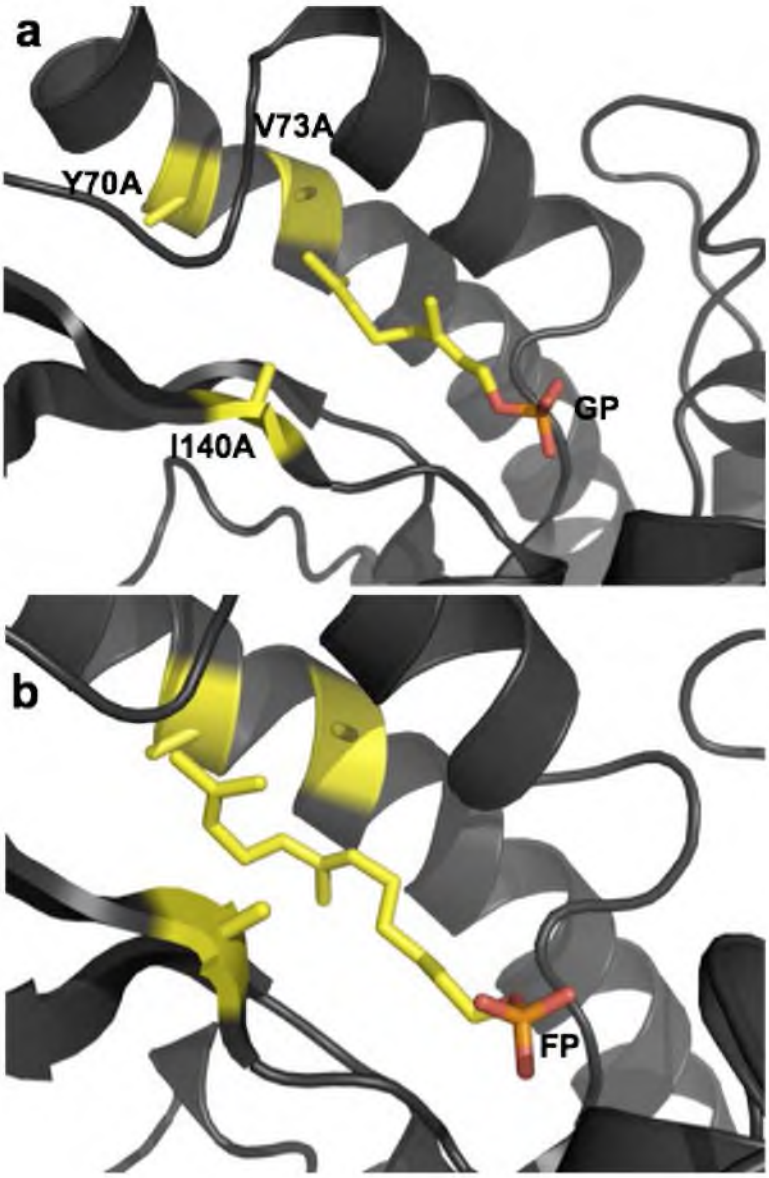


Figure 4.5. Modeled GP (a) and FP (b) molecule in the Mutant X active site.

The residual IP kinase activity of Mutant X is slightly higher than that of Mutant 12 (k_{cat} = 1.0 × 10^{-2} s^{-1}, K_M = 7.9 (± 1.2) × 10^3 μM), with a catalytic efficiency, k_{cat}/K_M = 1.3 M^{-1} s^{-1} that is 1.3 × 10^6 –fold lower than wild type THA IPK. The Michaelis-Menten curve for the phosphorylation of IP by Mutant X is found in the Appendix.

Similar to the profile observed in Mutant 12, the mutations on the IP binding site did not affect the kinetic parameters for the use of ATP in the presence of the three different substrates. The k_{cat}s and K_Ms for the use of ATP by Mutant X are similar to those measured for Mutant 12 and native THA IPK, and the resulting catalytic efficiencies are similar to those of native THA IPK with an average of 1.3 $M^{-1}s^{-1}$ (Table 4.2). The Michaelis-Menten curves for the use of ATP by Mutant X are found in the Appendix.

The formation of GPP and FPP upon incubation of Mutant X with synthesized substrates GP and FP was detected by UPLC and MS (see Appendix for MX GPP and MX FPP). A peak with retention time of 1.19 min on a C18 column, and a mass of 313, signals the presence of the negative ion form of GPP ($C_{10}H_{19}O_7P_2^-$). This product is well separated from the GP substrate peak (retention time 3.81 min, mass 233 for the negative ion ($C_{10}H_{18}O_4P^-$). Moreover, a small peak with a retention time of 4.83 min and a mass of 381 confirms the formation of FPP (mass 381, $C_{15}H_{27}O_7P_2^-$), separated from the starting substrate FP with a retention time of 5.43 min and mass of 301 ($C_{15}H_{26}O_4P^-$). Finally, using a C4 column, the separation of IPP from IP was unsuccessful, although both species were detected in the void volume (retention times less than 1 min) with masses of 245 and 165 for the negative ion species $C_5H_{11}O_7P_2^-$ and $C_5H_{10}O_3P^-$, respectively (see Appendix).

Mutant 7 (Tyr70Ala, Val130Ala and Ile140Ala). Among all four mutants tested, Mutant 7 has the highest turnover number, k_{cat}^{GP}= 10.1 (± 0.7) s^{-1}, similar to k_{cat}^{IP} for wild-type THA IPK (8.0 s^{-1}) and more than twice the activity of Mutant X. The k_{cat}^{GP} for Mutant 7 is also more than 200-fold greater than the promiscuous activity of wild-type THA IPK for GP. K_M^{GP} = 8.0 × 10^3 μM for Mutant 7 is slightly higher than for Mutant 12, Mutant X, and wild-type THA IPK, leading to a catalytic efficiency of 1.3 × 10^3 $M^{-1}s^{-1}$ that is only 10^3-fold lower than the catalytic efficiency of wild-type THA IPK for its native substrate, IP. The turnover number and Michaelis constant for FP by Mutant 7 is similar to that of Mutant X, with a k_{cat}^{FP} = 1.5 (± 0.1) s^{-1} and K_M^{FP} of 1.5 (± 0.2) × 10^3 μM, leading to k_{cat}/K_M = 9.9 × 10^2. Finally, the remnant IP kinase activity is also similar to those of the previous mutants, with low k_{cat}^{IP} = 4.8 × 10^{-1} s^{-1} that is almost 2000-fold lower than wild-type IPK. K_M^{IP} for Mutant 7 is also significantly higher (2000-fold) than that of wild-type THA IPK and similar to those measured for Mutant 12 and Mutant X. Figures 4.6a and 4.6b show Mutant 7 with bound GP and FP in the active site. The Michaelis-Menten curves for the kinase activities of Mutant 7 for GP, FP and IP are found in the Appendix. Unlike Mutant 12 and Mutant X, Mutant 7 was not saturated at 12.5 mM GP. At GP concentrations of 12.5 mM or higher, precipitation was observed within 7.5 minutes of initiation with enzyme, making accurate initial rates impossible to obtain. Thus, the fitted Michaelis-Menten curve for Mutant 7 shows only initial rate points corresponding to GP concentrations less than or equal to 12.5 mM (Appendix). Since precipitation of FP was observed at concentrations higher than 9.5 mM, initial rate measurements were carried out at this FP concentration or lower. However, unlike when GP is used as a substrate, Mutant 7 reached saturation with FP before precipitation was

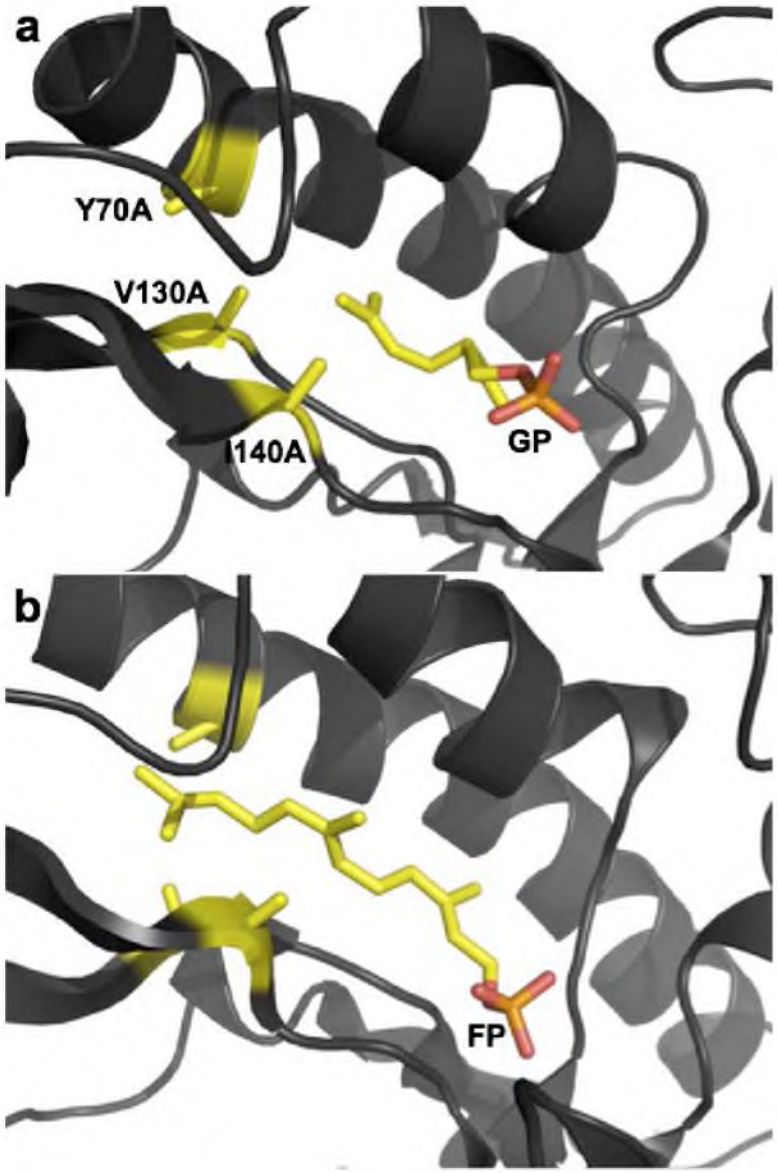


Figure 4.6. Modeled GP (a) and FP (b) molecules in the active site of Mutant 7.

observed. Given these observations, the measurement of kinetic parameters for the use of ATP by Mutant 7 were performed at the highest possible GP (10 mM) and FP (7 mM) concentrations before precipitation was observed. Nevertheless, the kinetic parameters k_{cat}^{ATP} and K_M^{ATP} for Mutant 7 remained consistent with those measured for wild-type THA IPK, Mutant 12 and Mutant X, with an average k_{cat}/K_M 1.0×10^6 $M^{-1}s^{-1}$. The Michaelis-Menten curves for use of ATP by Mutant 7 are found in the Appendix.

GPP produced by Mutant 7 was confirmed by UPLC-MS, corresponding to a single sharp peak with a retention time of 1.11 min and mass of 313 for the negative ion species, $C_{10}H_{19}O_7P_2^-$. A much smaller peak corresponding to the substrate GP ($C_{10}H_{18}O_4P^-$) was also found with a retention time of 3.37 min and mass of 233. The production of FPP by Mutant 7 was also supported by the presence of a small peak with a retention time of 4. 70 min and mass of 381 ($C_{15}H_{27}O_7P_2^-$), separated from the strong FP peak with a retention time of 5.45 min and mass of 301 ($C_{15}H_{26}O_4P^-$). Finally, neither IPP nor IP was purified from each other using the C4 column, although masses for the negative ions species for both compounds were found at 245 ($C_5H_{11}O_7P_2^-$) and 165 ($C_5H_{10}O_3P^-$) (see Appendix).

Mutant 11 (Tyr70Ala, Val130Ala, Ile140Ala). Among the four mutants, Mutant 11 did not contain the Tyr70Ala mutation, resulting in a more compact substrate-binding site that was presumed to either not bind FP or do so with FP in a more compact conformation. However, this minimal mutation is sufficient to bind a fully extended GP molecule (Figure 4.7a). Product assays using γ-[^{32}P] ATP showed that the FP kinase activity of Mutant 11 was undetectable (Figure 4.2f). Thus, only the GP kinase activity of Mutant 11 was measured, in conjunction with the residual IP kinase activity. In addition,

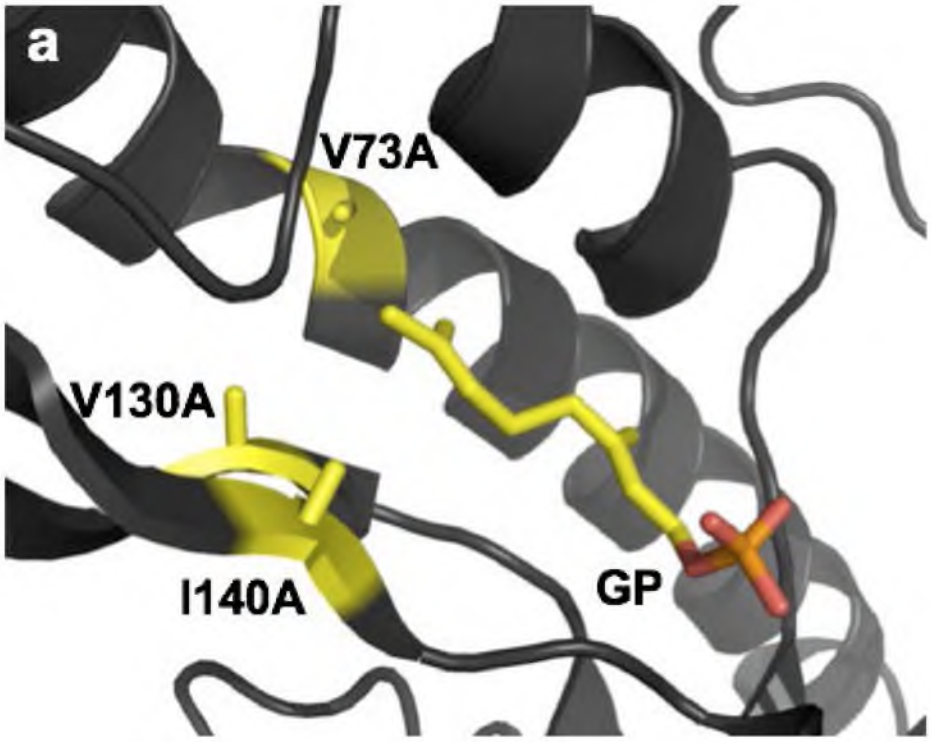


Figure 4.7. Modeled GP molecule in the active site of Mutant 11.

as in the case of Mutant 7, the GP kinase reaction did not reach saturation before precipitation was observed, so that initial rate measurements for the steady-state kinetic investigations of this mutant were performed below 12.5 mM. The k_{cat}^{GP} = 2.8 (± 0.4) s^{-1} for Mutant 11 reflects a 60-fold improvement to the turnover number of wild-type THA IPK. K_M^{GP} = 9.5 (± 0.3) × 10^3 μM for Mutant 7 is the highest among all four mutants, leading to a catalytic efficiency for the phosphorylation of GP, k_{cat}/K_M = 3.0 × 10^2 $M^{-1}s^{-1}$ that is 30-fold larger than the catalytic efficiency of the wild-type THA IPK using GP. This translates to more than a 6000-fold decrease in catalytic efficiency compared to the native activity of THA IPK for its substrate IP. The Michaelis-Menten curve for the GP kinase reaction catalyzed by Mutant 11 is found in the Appendix.

The residual IP kinase activity of Mutant 11 is comparable to those of the other mutants, k_{cat}^{IP} = 1.1 (± 0.1) × 10^{-2} s^{-1} and almost a thousand-fold lower than that of wild-type THA IPK. K_M^{IP} = 4.2 (± 0.7) × 10^3 μM is also significantly higher than that of wild-type THA IPK. Thus, the catalytic efficiency for the phosphoryl transfer to IP catalyzed by Mutant 11 is 2.6 $M^{-1}s^{-1}$, more than 70,000-fold less than that for native THA IPK. The Michaelis-Menten curve for the IP kinase reaction catalyzed by Mutant 11 is found in the Appendix.

The kinetic constants for the utilization of ATP in the kinase reaction by Mutant 11 are very similar to those of other mutants as well as that of wild-type THA IPK, regardless of whether GP or IP was used as a cosubstrate. k_{cat}^{ATP} = 1.3 (± 0.1) s^{-1} for GP and 1.4 (± 0.1) s^{-1} for IP. K_M^{ATP} = 1.2 (± 0.2) s^{-1} and 1.9 (±0.5) s^{-1}, for GP and IP as cosubstrates, respectively. The calculated catalytic efficiencies, k_{cat}/K_M, are very close to that of wild-type THA IPK, at 1.06 × 10^6 $M^{-1}s^{-1}$ (GP) and 7.9 × 10^5 $M^{-1}s^{-1}$ (IP). The

Michaelis-Menten curves for the utilization of ATP by Mutant 11 are shown in the Appendix.

Finally, as for the other mutants, GPP production by Mutant 11 was confirmed by a sharp peak with a retention time of 1.11 min, corresponding to a mass of 313 ($C_{10}H_{19}O_7P_2^-$). A smaller peak with a retention time of 3.48 min, corresponding to GP was also found, with a corresponding mass of 233 ($C_{10}H_{18}O_4P^-$) (see Appendix, M11 GPP). Although separation of IPP and IP was not achieved using a C4 column, masses each corresponding to their negative ion species were found in the void volume: 245 for IPP ($C_5H_{11}O_7P_2^-$) and 165 for IP ($C_5H_{10}O_3P^-$) (See Appendix, M11 IPP).

Discussion

A total of 8 THA IPK mutants were created by the structure-based redesign of the THA IPK active site, with mutations at locations occupied by bulky amino acids in the IP binding site presumably influencing chain length specificity of this enzyme *(40)*. Our goal was to create IPK variants that possess the ability to phosphorylate isoprenoid monophosphates longer than IP such as GP and FP. The mutants would be useful for synthesizing radiolabeled isoprenoid diphosphates with ^{32}P in the β-position. These labeled compounds have many potential applications, including assays where isoprenoid diphosphates are substrates. In addition, these kinases can be used *in vivo* to recycle isoprenoid monophosphates from degradation of farnesyl diphosphate *(41,51-52)*.

The redesign of IPK relied on the availability of an enzyme structure at good resolution. We were able to determine the structures of MTH and THA IPKs at 2.0 Å *(40)*. Kinetic characterization of both enzymes showed that the THA IPK homologs have

comparable catalytic turnovers, with THA IPK being slightly better *(53)*. With this in mind, we proceeded to create THA IPK mutants and characterized the kinetic behavior of those with the most prominent GP and FP kinase activities.

For all mutants, K_M^{IP} increased 10^3-fold while k_{cat}^{IP} decreased 10^3-fold. The increase in K_M^{IP} might result from a corresponding increase in the size of the IP binding site that allows IP to bind in a variety of unproductive conformations. This in turn might cause the improper alignment of the terminal phosphate of IP relative to ATP, as well as with the catalytic residues in the IP binding site. In contrast, modest changes in K_M^{GP} were observed in the THA IPK mutants relative to that of wild type THA IPK, in conjunction with a 23- to 214-fold increase in the turnover number of GP. Thus, it appears that the expansion of substrate binding site allowed GP to orient better in the active site, leading to an increase in the productive catalysis for the mutant enzymes. Among the enzymes studied, Mutant 7, which contains the Tyr70Ala, Val130Ala and Ile140Ala mutations, has a k_{cat}^{GP} = 10.1 s^{-1} that is comparable to k_{cat}^{IP} = 8.0 s^{-1} of wild-type THA IPK *(53)*. This mutant represents a good starting point for the development of a better "geranyl phosphate kinase" by means of further mutations to optimize the binding site for GP. The other mutants have comparable GP kinase activities, and the results of our product assays (autoradiography and liquid chromatography) showed almost complete conversion of substrates to products when 2 mM of GP and 5 mM of ATP were used, suggesting an equilibrium that favors GPP formation. Crystal structures of these mutants in complex with GP will help pinpoint additional mutations that can make the geranyl phosphate kinase activity more efficient by lowering K_M^{GP} through better binding of the substrate. It appears that the three mutants with the highest GP kinase activities

(Mutant 7, Mutant X, and Mutant 11) have smaller substrate binding pockets than Mutant 12, which seems to suggest that a bigger cavity does not necessarily translate to higher activity. Rather, the GP binding site must be just wide enough to accommodate a GP molecule while at the same time restricting it to one or a few binding poses suitable for catalysis.

We have also shown that the mutants containing the Tyr70Ala mutation (Mutant 12, Mutant 7 and Mutant 11) have secondary FP kinase activities that are lower than the observed GP kinase activities, supported qualitatively by product assays and quantitatively by kinetic measurements. Our modeling experiments showed that the Tyr70Ala mutation is necessary to bind a fully extended FP molecule in addition to mutations in the front of the IP binding site. However, FP can bind in several different different orientations through rotation around single bonds in the isoprenoid chain. Thus, although in autoradiography assays we did not detect FP kinase activity for Mutant 11, which does not contain the Tyr70Ala mutation, it is likely that this mutant also possesses, albeit weak, FP kinase activity that may become detectable under forcing conditions. A mutant of MJ IPK with similar mutations has FP kinase activity detected by the coupled IPK-sesquiterpene synthase assay, in which the formation of FPP by mutant IPK is coupled with tobacco 5-epi aristolochene synthase activity to form 5-epi aristolochene *(41,54)*.

Our steady-state measurements showed that Mutant 12, Mutant X and Mutant 7 have K_M^{FP}s that are lower than K_M^{GP}, which might suggest that a molecule of FP can bind better through more van der Waals interactions between its longer isoprenoid tail and the walls of the mutant active site. However, it might be oriented less optimally in the active

site than a GP molecule due to its ability to interconvert among different conformations, leading to a lower k_{cat}. The low FP kinase activities of these mutants are more than a hundred-fold higher than their residual IP kinase activities but at most 10-fold lower than their GP kinase activities. Thus, it appears that the mutants possess triple isoprenoid monophosphate activities, with GP kinase activity being the strongest among the three.

We have also observed that kinetic parameters for the utilization of ATP by the mutants are very similar to those of wild-type THA IPK, regardless of the identity of the cosubstrate. The catalytic efficiency of using ATP as a phosphoryl group donor is in the order of 10^6 $M^{-1}s^{-1}$ in the mutants as well as in the native enzyme. This is expected since the mutations were confined to the IP binding site, and there are no extensive interactions between the substrates aside from their terminal phosphate groups. There appears to be no cooperative substrate specificity as the N-terminal lobe, which binds the isoprenoid substrate, and the C-terminal lobe, which binds ATP, act independently of each other. This phenomenon has been observed in the conversion of human pFTase to pGGTase I, where the kinetic parameters for the Ca_1a_2X substrates of pFTase were not altered in the presence of change in substrate specificity from FPP to GGPP *(39)*.

We have tested eight THA IPK mutants and further characterized four variants and observed significant improvements in GP kinase activities at the expense of native IP kinase activity. We have also observed FP kinase activities in three mutants that are inferior to their GP kinase activities. Thus, a directed shift in reaction specificity from IPK to GPK has been accomplished, and to our knowledge, this is the first instance in which an artificial GPK was reported and partially characterized. Dellas and Noel reported the conversion of a thermophilic IPK homolog to FPK, although the activity

toward FP was not established *(41)*. Other groups have reported the presence of isoprenol kinases able to catalyze the successive phosphorylation of farnesol and geranylgeraniol using ATP, GTP, UTP and CTP, to form FPP and GGPP in a probable salvage pathway for recycling isoprenoid alcohols *(43-46,55)*. Despite this accomplishment, there exist many other possible IPK variants that can bind different conformations of GP and FP that need to be identified, and where the high-throughput coupled kinase/terpene synthase assay suggested by Dellas and Noel would be ideal for rapid and qualitative analysis. In addition, the mutations performed in this work were X to Ala mutations. It would be interesting to see what other types and combinations of mutations would yield engineered IPKs with the highest isoprenoid monophosphate kinase activities. Finally, similar mutational studies on mesophilic homologs of IPK could exhibit improvements in catalytic properties of engineered enzymes targeted for characterization and practical application at ambient temperatures.

References

1. Chen, R. (2001) Enzyme engineering: rational redesign versus directed evolution, *Trends Biotechnol. 19*, 13-14.

2. Ohnuma, S., Nakazawa, T., Hemmi, H., Hallberg, A., Koyama, T., Ogura, K., and Nishino, T. (1996) Conversion from farnesyl diphosphate synthase to geranylgeranyl diphosphate synthase by random chemical mutagenesis, *J. Biol. Chem. 271*, 10087-10095.

3. Ohnuma, S., Hirooka, K., Hemmi, H., Ishida, C., Ohto, C., and Nishino, T. (1996) Conversion of product specificity of archaebacterial geranylgeranyl diphosphate synthase, *J. Biol. Chem. 271*, 18831-18837.

4. Chen, A., Kroon, P. A., and Poulter, C. D. (1994) Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure, *Protein Sci. 3*, 600-607.

5. Clarke, C. F., Tanaka, R. D., Svenson, K., Wamsley, M. Fogelman, A. M., and Edwards, P. A. (1987) Molecular cloning and sequence of a cholesterol-repressible enzyme related to prenyltransferase in the isoprene biosynthetic pathway, *Mol. Cell. Biol. 7*, 3138-3146.

6. Anderson, M. S., Yarger, J. G., Burck, C. L., and Poulter, C. D. (1989) Farnesyl diphosphate synthetase: molecular cloning, sequence and expression of an essential gene from *Saccharomyces cerevisiae*, *J. Biol. Chem. 264*, 19176-19184.

7. Fujisaki, S., Hara, H., Nishimura, Y., Horiuchi, K., and Nishino, T. (1990) Cloning and nucleotide sequence of the ispA gene responsible for farnesyl diphosphate synthase activity in *Escherichia coli*, *J. Biochem. (Tokyo) 108*, 995-1000.

8. Koyama, T., Obata, S., Osabe, M., Takeshita, A., Yokoyama, K., Uchida, M., Nishino, T., and Ogura, K. (1993) Thermostable farnesyl diphosphate synthase of *Bacillus stearothermophilus*: molecular cloning, sequence determination, overproduction, and purification, *J. Biochem. (Tokyo) 113*, 355-363.

9. Wilkin, D. J., Kutsunai, S. Y., and Edwards, P. A. (1990) Isolation and sequence of the human farnesyl pyrophosphate synthetase cDNA. Coordinate regulation of the mRNAs for farnesyl pyrophosphate synthetase, 3-hydroxy-3-methylglutaryl coenzyme A reductase, and 3-hydroxy-3-methylglutaryl coenzyme A synthase by phorbol ester. *J. Biol. Chem. 265*, 4607-4614.

10. Math, S. K., Hearst, J. E., and Poulter, C. D. (1992) The crtE gene in *Erwinia herbicola* encodes geranylgeranyl diphosphate synthase, *Proc. Natl. Acad. Sci. U. S. A.* 89, 6761-6764.

11. Carattoli, A., Romano, N., Ballario, P., Morelli, G., and Macino, G. (1991) The *Neurospora crassa* carotenoid biosynthetic gene (albino 3) reveals highly conserved regions among prenyltransferases, *J. Biol. Chem. 266*, 5854-5859.

12. Ohnuma, S., Suzuki, M., and Nishino, T. (1994) Archaebacterial ether-linked lipid biosynthetic gene. Expression cloning, sequencing, and characterization of geranylgeranyl-diphosphate synthase, *J. Biol. Chem. 269*, 14792-14797.

13. Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., and Harashima, K. (1990) Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*, *J. Bacteriol. 172*, 6704-6712.

14. Ashby, M. N., and Edwards, P. A. (1990) Elucidation of the deficiency in two yeast coenzyme Q mutants. Characterization of the structural gene encoding hexaprenyl pyrophosphate synthetase. *J. Biol. Chem. 265*, 13157-13164.

15. Koike-Takeshita, A., Koyama, T., Obata, S., and Ogura, K. (1995) Molecular cloning and nucleotide sequences of the genes for two essential proteins constituting a novel enzyme system for heptaprenyl diphosphate synthesis. *J. Biol. Chem. 270*, 18396-18400.

16. Asai, K., Fujisaki, S., Nishimura, Y., Nishino, T., Okada, K., Nakagawa, T., Kawamukai, M., and Matsuda. H. (1994) The identification of *Escherichia coli* ispB (cel) gene encoding the octaprenyl diphosphate synthase, *Biochem. Biophys. Res. Commun. 202*, 340-345.

17. Ohnuma, S., Hirooka, K., Ohto, C. and Nishino, T. (1997) Conversion from archaeal geranylgeranyl diphosphate synthase to farnesyl diphosphate synthase, *J. Biol. Chem. 272*, 5192-5198.

18. Tarshis, L. C., Yan, M., Poulter, C. D., and Sacchetini, J. C. (1994) *Biochemistry 33*, 10871-10877.

19. Stanley Fernandez, S. M., Kellogg, B., and Poulter, C. D. (2000) Farnesyl diphosphate synthase. Altering the catalytic site to select for geranyl diphosphate activity, *Biochemistry 39*, 15316-15321.

20. Casey, P. J., and Seabra, M. C. (1996) Protein prenyltransferases. *J. Biol. Chem. 271*, 5289-5292.

21. Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2004) Crystallgraphic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity, *J. Mol. Biol. 343*, 417-433.

22. Spence, R. A., and Casey, P. J. (2001) The Enzymes (Tamanoi, F., and Sigman, D. S., Eds.) pp 1-18, Academic Press, San Diego, CA.

23. Brown, M. S., Goldstein, J. L., Paris, K. J., Burnier, J. P., and Masters, J. C. (1992) Tetrapeptide inhibitors of protein farnesyltransferase: amino terminal substitutions in phenylalanine-containing tetrapeptides restores farnesylation, *Proc. Natl. Acad. Sci. U. S. A. 89*, 8313-8316.

24. Winter-Vann, A. M., and Casey, P. J. (2005) Post-prenylation-processing enzymes as new targets in oncogenesis, *Nat. Rev. Cancer 5*, 405-412.

25. Silvius, J. R., and l'Heureux, F. (1994) Fluorimetric evaluation of the affinities of isoprenylated peptides for lipid bilayers, *Biochemistry 33*, 3014-3022.

26. Marshall, C. J. (1993) Protein prenylation: a mediator of protein-protein interactions, *Science 259*, 1865-1866.

27. Gelb, M. H. (1997) Protein prenylation, et cetera: signal transduction in two dimensions, *Science 275*, 1750-1751.

28. Yokoyama, K., Zimmerman, K., Scholten, J., and Gelb, M. H. (1997) Differential prenyl pyrophosphate binding to mammalian protein geranylgeranyltransferase-I and protein farnesyltransferase and its consequences on the specificity of protein prenylation, *J. Biol. Chem. 272*, 3944-3952.

29. Reiss, Y., Brown, M. S., and Goldstein, J. L. (1992) Divalent cation and prenyl pyrophosphate specificities of the protein farnesyltransferase from the rat brain, a zince metalloenzyme, *J. Biol. Chem. 267*, 6403-6408.

30. Long, S. B., Casey, P. J., and Beese, L. S. (1998) Co-crystal structure of protein farnesyltransferase with a farnesyl diphosphate substrate, *Biochemistry 37*, 9612-9618.

31. Long, S. B., Casey, P. J., and Beese, L. S. (2000) The basis for K-Ras4B binding specificity to protein farnesyltransferase revealed by a 2 Å resolution ternary complex structures, *Structure 8*, 209-222.

32. Long, S. B., Hancock, P. J., Kral, A. M., Hellinga, H. W., and Beese, L. S. (2001) The crystal structure of human farnesyltransferase reveals the basis for inhibition by CaaX tetrapeptides and their mimetics, *Proc. Natl. Acad. Sci. U. S. A. 98*, 12948-12953.

33. Long, S. B., Casey, P. J., and Beese, L. S. (2002) Reaction path of protein farnesyltransferase at atomic resolution, *Nature 419*, 645-650.

34. Reid, T. S., and Beese, L. S. (2004) Crystal structures of the anticancer clinical candidates R115777 (tipifarnib) and BMS-214662 complexed with protein farnesyltrasnferase suggest a mechanism of FTI selectivity, *Biochemistry 43*, 6877-6884.

35. de Solms, S. J., Ciccarone, T. M., MacTough, S. C., Shaw, A. W., Buser, C. A., Ellis-Hutchings, M., Fernandes, C., Hamilton, K. A., Huber, H. A., Kohl, N. E., Lobell, R. B., Robinson, R. G., Tsou, N. N., Walsh, E. S., Graham, S. L., Beese, L. S., and Taylor, J. S. (2003) Dual protein farnesyltransferase-geranylgeranyltransferase I inhibitors as potential cancer therapeutic agents, *J. Med. Chem. 46*, 2973-2984.

36. Reid, T. S., Long, S. B., and Beese, L. S. (2004) Crystallographic analysis reveals that anticancer clinical candidate L-778, 123 inhibits protein farnesyltransferase and geranylgeranyltransferase-I by different binding modes, *Biochemistry 43*, 9000-9008.

37. Strickland, C. L, Windsor, W. T., Syto, R., Wang, L., Bond, R., Wu, Z., Schwartz, J., Le, H.V., Beese, L. S., and Weber, P. C. (1998) Crystal structure of farnesyl protein transferase complexed with a CaaX peptide and farnesyl diphosphate analogue, *Biochemistry 37*, 16601-16611.

38. Taylor, J. S., Reid, T. S., Terry, K. L., Casey, P. J., and Beese, L. S. (2003) Structure of mammalian protein geranylgeranyltransferase type-I, *EMBO J.* 22, 5963-5974.

39. Terry, K. L., Casey, P. J., and Beese, L. S. (2006) Conversion of protein farnesyl transferase to a geranylgeranyltransferase, *Biochemistry 45*, 9746-9755.

40. Mabanglo, M. F., Schubert, H. L. Chen, M., Hill, C. P., and Poulter, C. D. (2010) X-ray structures of isopentenyl phosphate kinase, *ACS Chem. Biol. 5(5),* 517-527.

41. Dellas, N., and Noel, J. P. (2010) Mutation of Archaeal isopentenyl phosphate kinase highlights mechanism and guides phosphorylation of additional isoprenoid monophosphates, *ACS Chem. Biol. 5(6)*, 589-601.

42. Thai, L., Rush, J. S., Maul, J. E., Devarenne, T., and Rodgers, D. L., Chappell, J., and Waechter, C. J. (1999) Farnesol is utilized for isoprenoid biosynthesis in plant cells via farnesyl pyrophosphate formed by successive monophosphorylation reactions, *Proc. Natl. Acad. Sci U. S. A. 96*, 13080-13085.

43. Inoue, H., Korenaga, T., Sagami, H., Koyama, T., and Ogura, K. (1994) Phosphorylation of farnesol by a cell-free system from *Botryococcus braunii*, *Biochem. Biophys. Res. Commun. 200*, 1036-1041.

44. Ohnuma, S., Watanabe, M., and Nishino, T. (1996) Identification and characterization of geranylgeraniol kinase and geranylgeranyl phosphate kinase from the Archaebacterium *Sulfolobus acidocaldarius*, *J. Biochem. (Tokyo) 119*, 541-547.

45. Westfall, D., Aboushadi, N., Shackelford, J. E., and Krisans, S. K. (1997) Metabolism of farnesol: phosphorylation of farnesol by rat liver microsomal and peroxisomal fractions, *Biochem. Biophys. Res. Commun. 230*, 562-568.

46. Bentinger, M., Grünler, J., Peterson, E., Swiezewska, E., and Dallner, G. (1998) Phosphorylation of farnesol in rat liver microsomes: properties of farnesol kinase and farnesyl phosphate kinase, *Arch. Biochem. Biophys. 353*, 191-198.

47. Keller, R. K., and Thompson, R. (1993) Rapid synthesis of isoprenoid diphosphate and their isolation in one step using either thin layer or flash chromatography, *J. Chromatography A 645*, 161-167.

48. Petterson, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt. D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera – a visualization system for exploratory research and analysis, *J. Comput Chem. 13*, 1605-1612.

49. Pilloff, D., Dabovic, K., Romanowski, M. J., Bonanno, J. B., Doherty, M., Burley, S. K., and Leyh, T. S. (2003) The kinetic mechanism of phosphomevalonate kinase, *J. Biol. Chem 278*, 4510-4515.

50. Leatherbarrow, R. J. (2001) GraFit Version 5, Erithacus Software Ltd., Horley, U. K.

51. Song, L. (2006) A soluble form of phosphatase in *Saccharomyces cerevisiae* capable of converting farnesyl diphosphate into *E,E*-farnesol, *Appl. Biochem. Biotechnol. 128*, 149-158.

52. Coleman, J. E. (1992) Structure and mechanism of alkaline phosphatase, *Annu. Rev. Biophys. Biomol. Struct. 21*, 441-483.

53. Chen, M., and Poulter, C. D. (2010) Characterization of thermophilic Archaeal isopentenyl phosphate kinase, *Biochemistry 49*, 207-217.

54. O'Maille, P E., Chappell, J., and Noel, J. P. (2004) A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases, *Anal. Biochem. 335*, 210-217.

55. Tachibana, A, Tanaka, T., Taniguchi, M., and Oi, S. (1996) Evidence for farnesol-mediated isoprenoid synthesis regulation in a halophilic archaeon, *Halopherax volcanii*, *FEBS Lett. 379*, 43-46.

APPENDIX

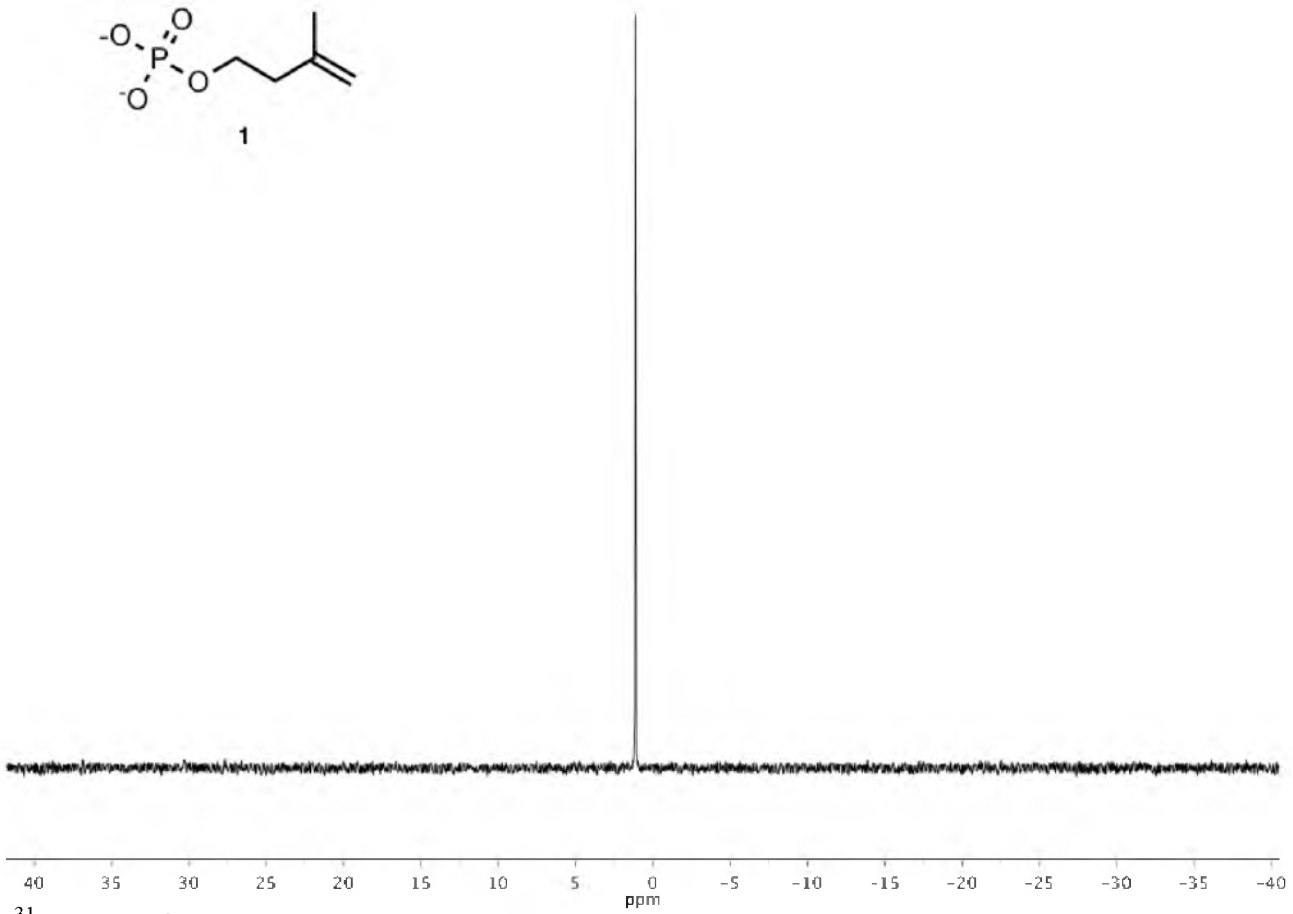
^{1}H and ^{31}P NMR Spectra of IP, GP and FP

MS/MS Fragmentation of IPK and FomA Product Fosfomycin Phosphate

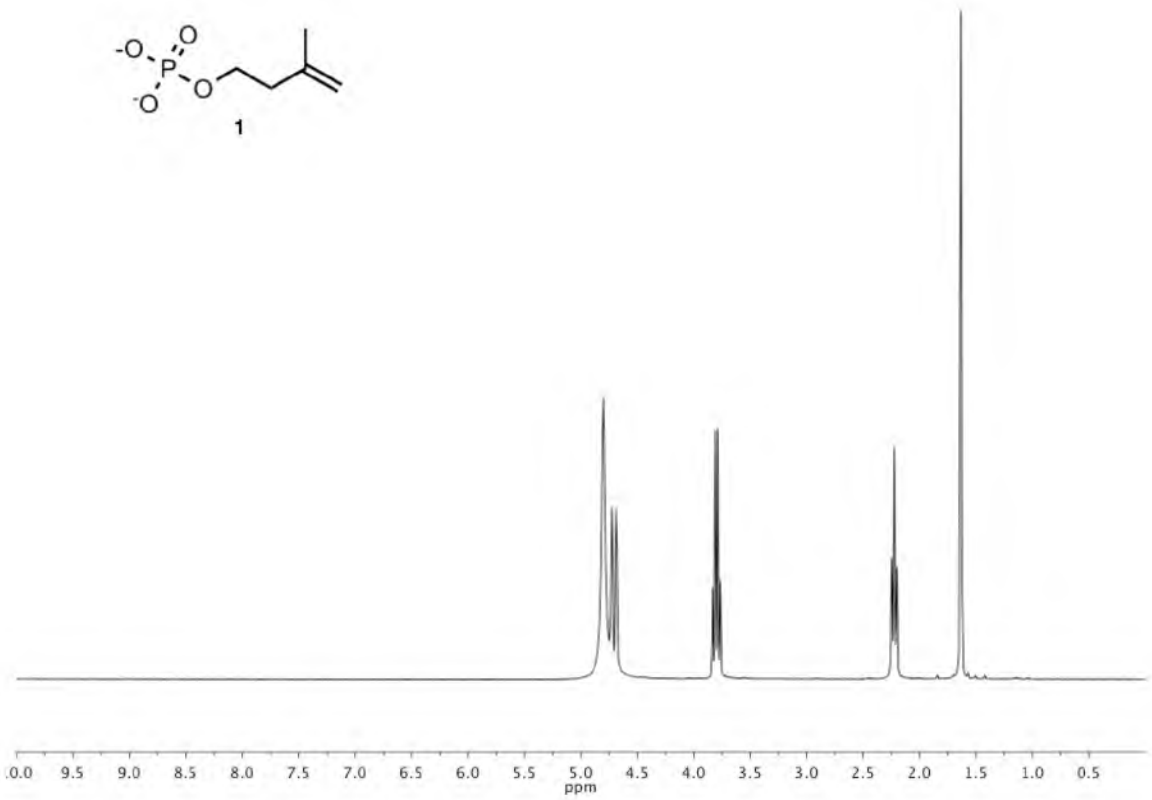
MS/MS Fragmentation of Fosfomycin Standard

UPLC- MS of Mutant 12, Mutant X, Mutant 7, and Mutant 11 Products

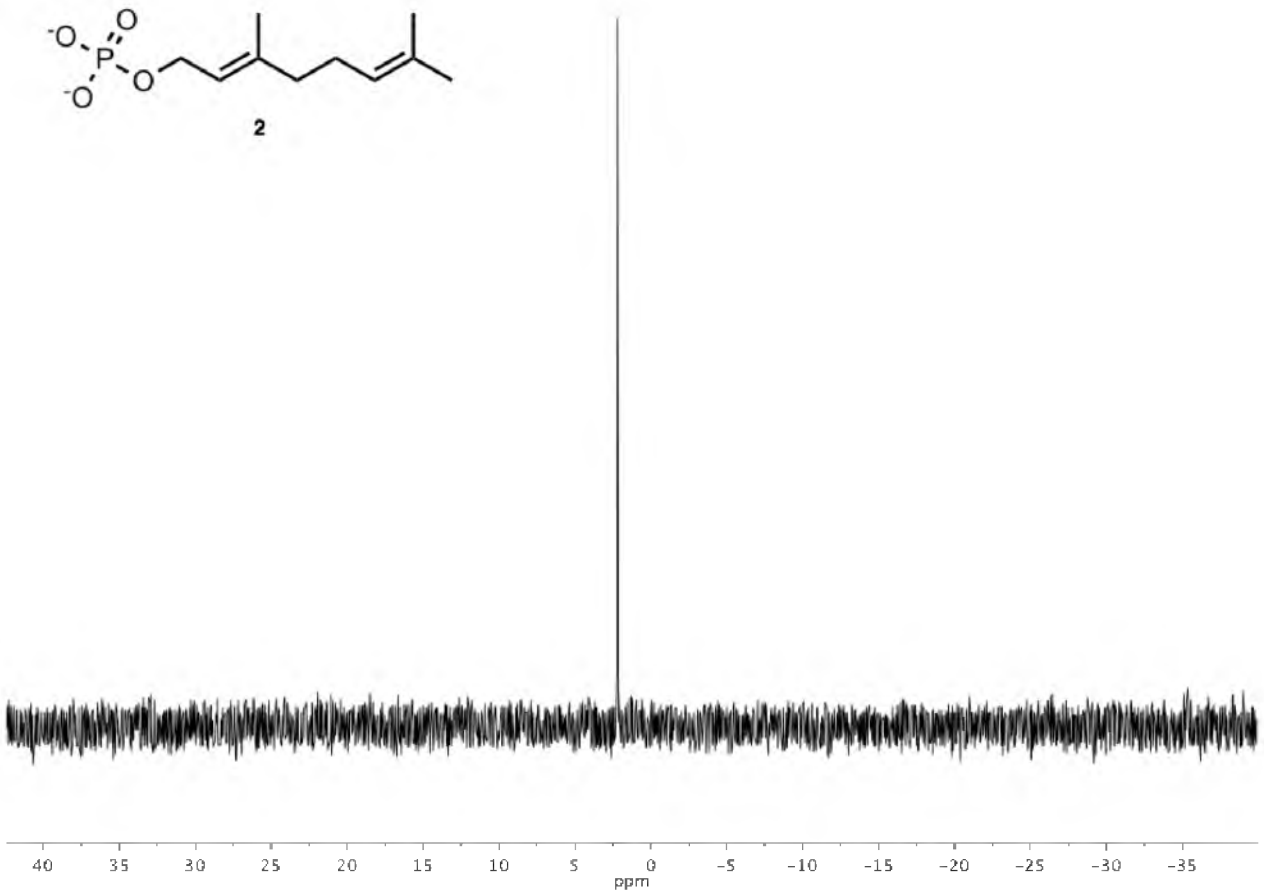
Michaelis-Menten Curves for Mutants 12, Mutant X, Mutant 7 and Mutant 11, using GP, FP, IP, and ATP as Substrates



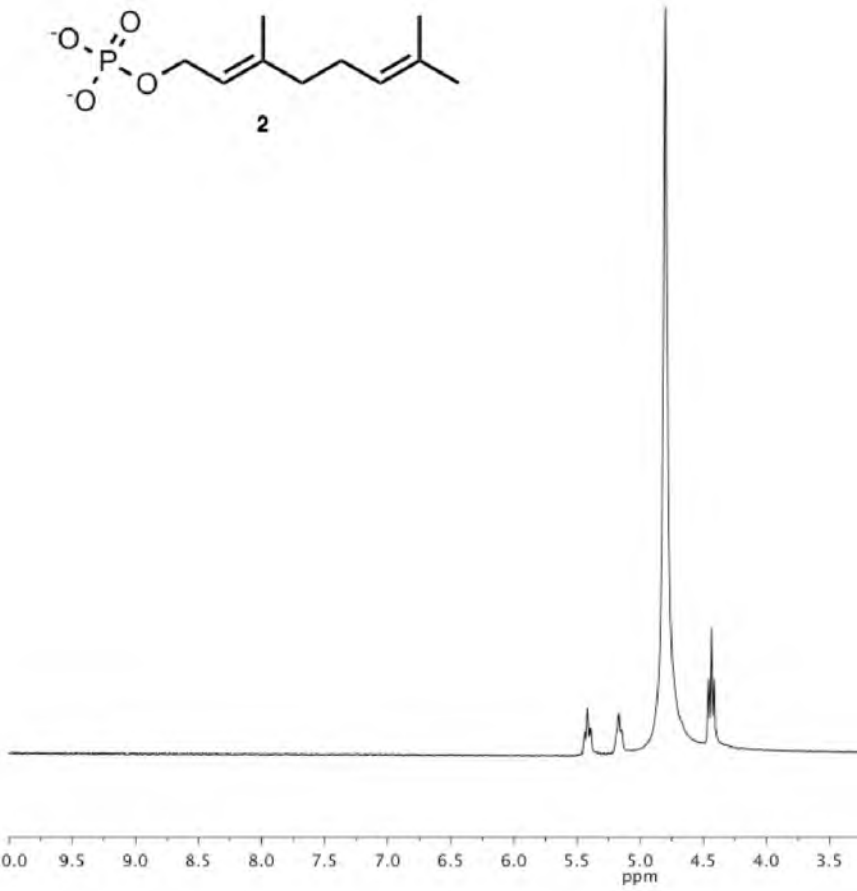
^{31}P NMR of IP



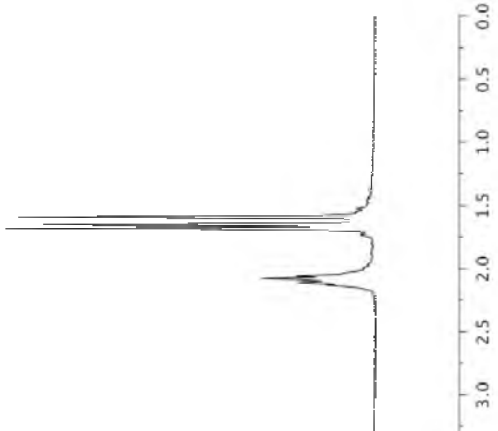
^{1}H NMR of IP

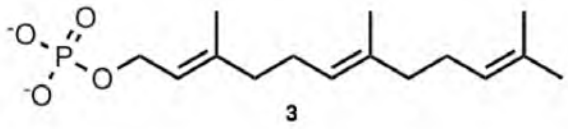


^{31}P NMR of GP

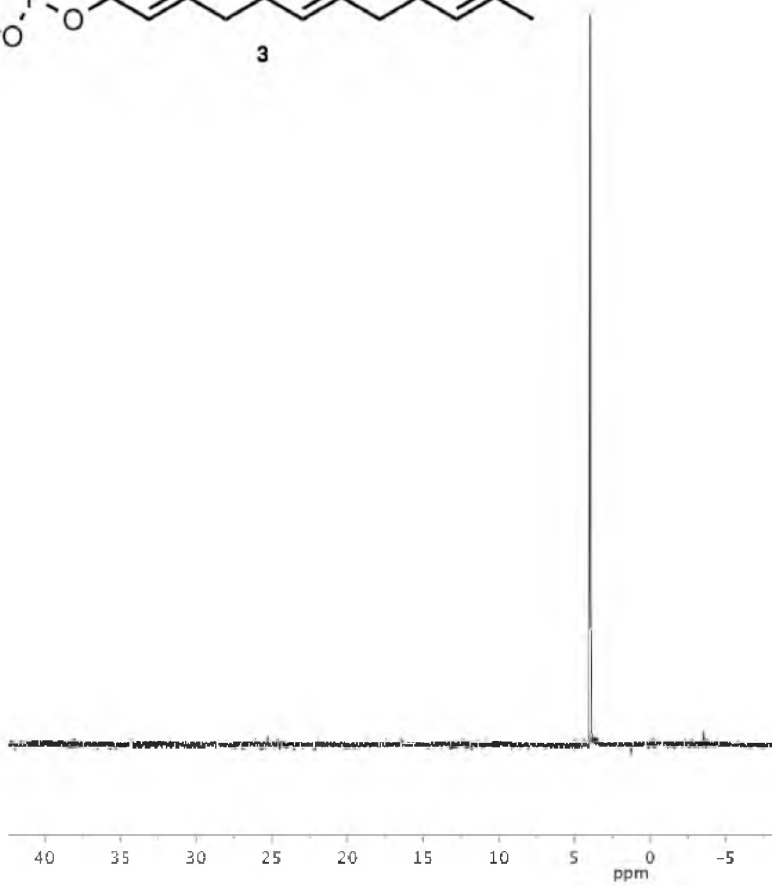


^{1}H NMR of GP


0.0
0.5
1.0
1.5
2.0
2.5
3.0

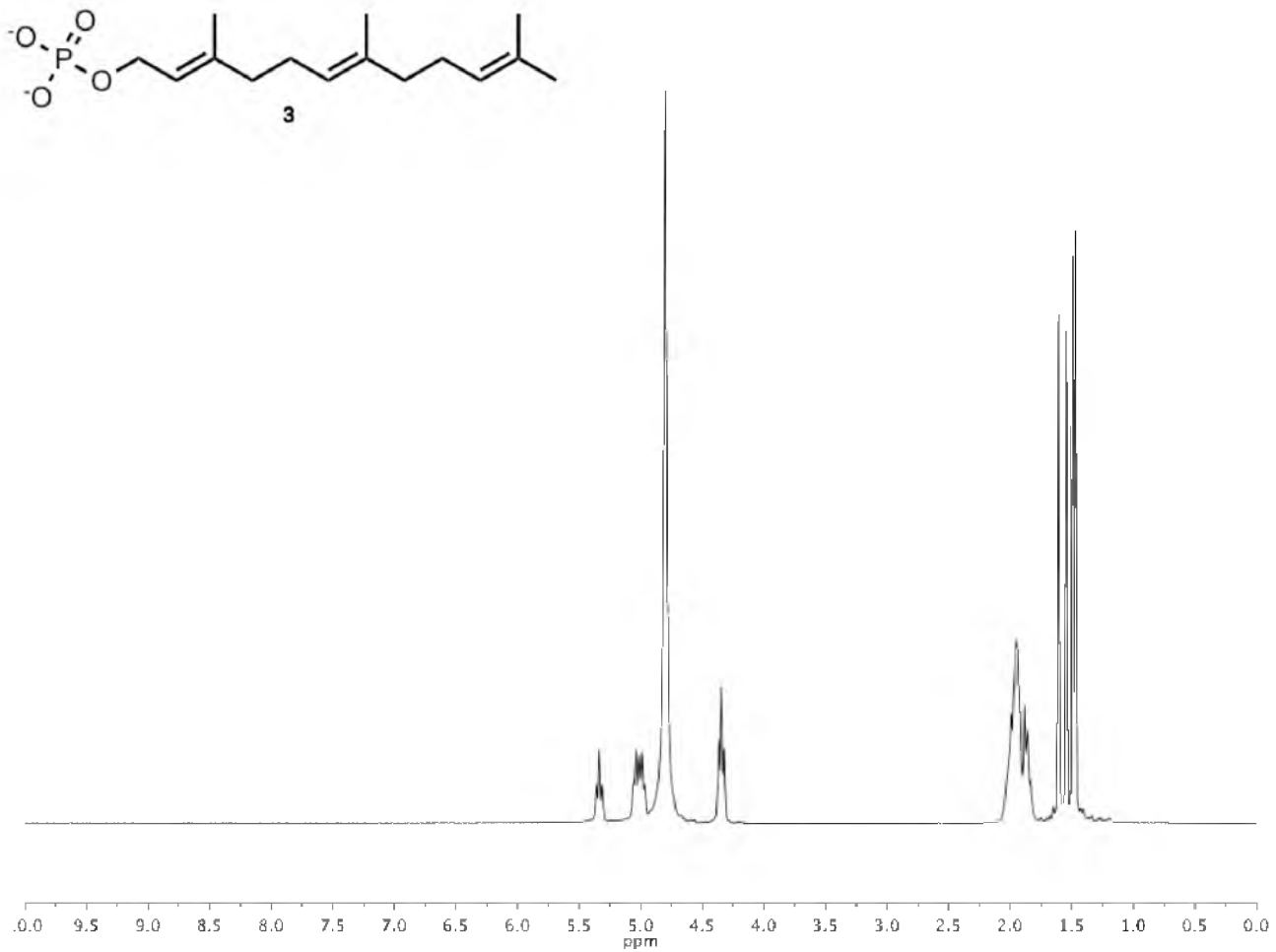


3

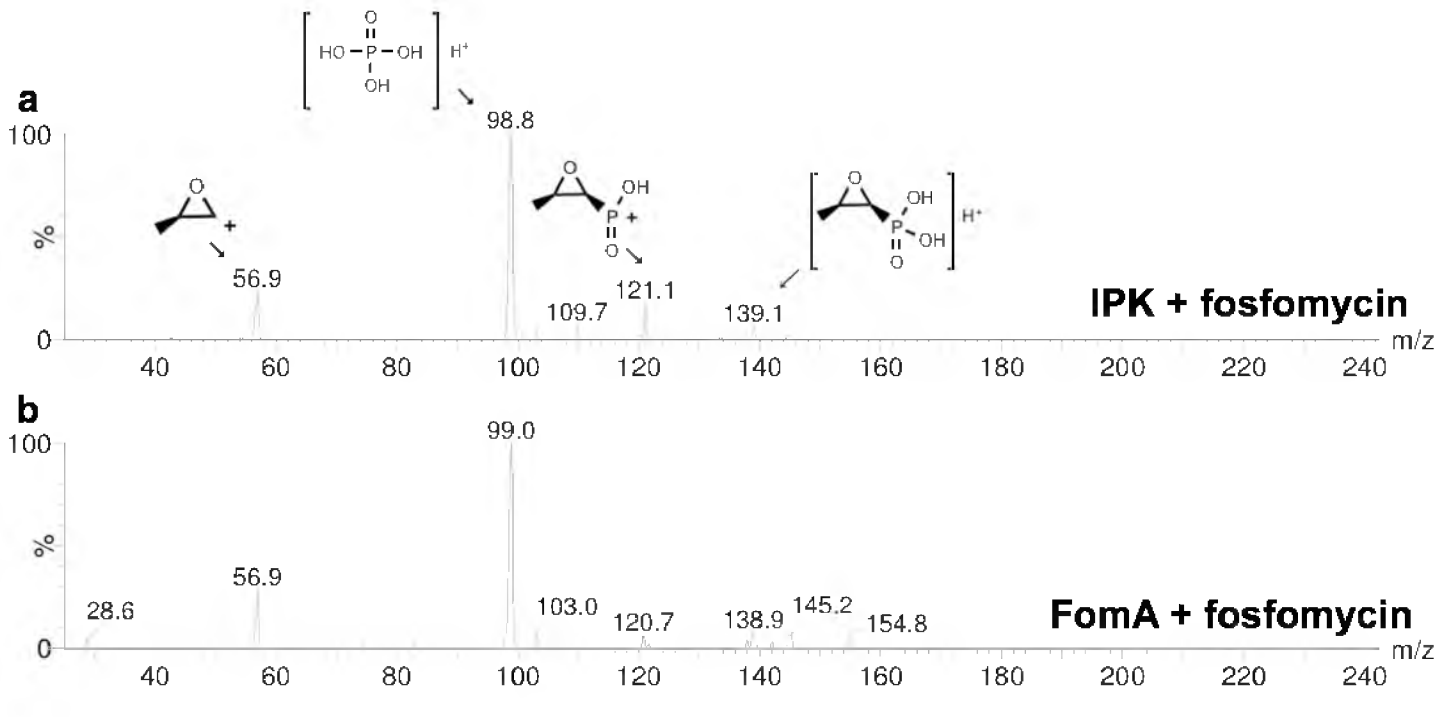


^{31}P NMR of FP

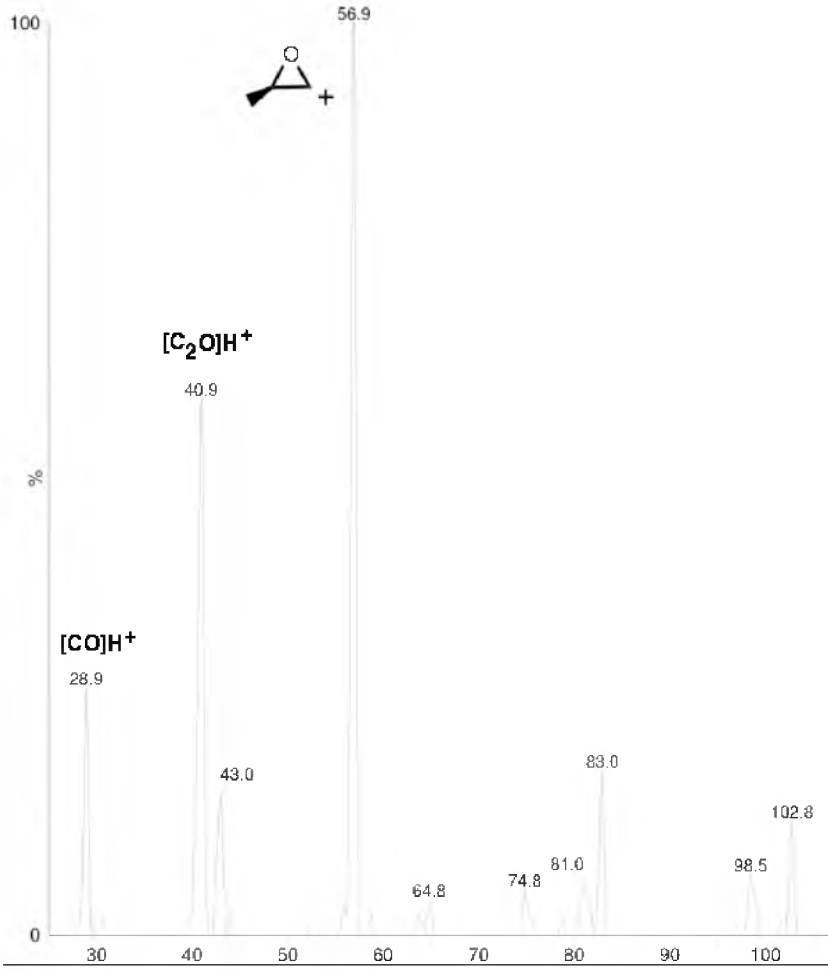
-10
-15
-20
-25
-30
-35



^{1}H NMR of FP

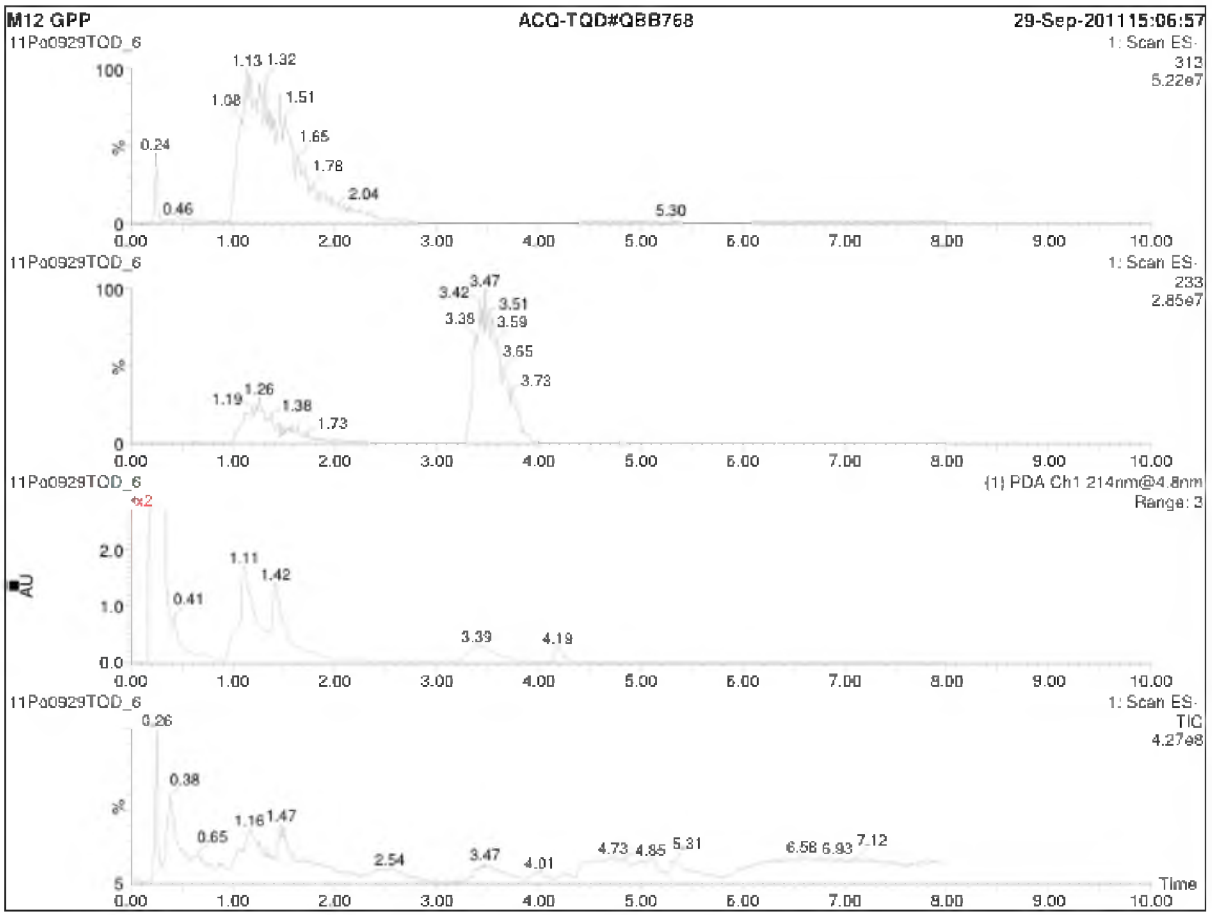


MS/MS Fragmentation of IPK and FomA Product Fosfomycin Phosphate

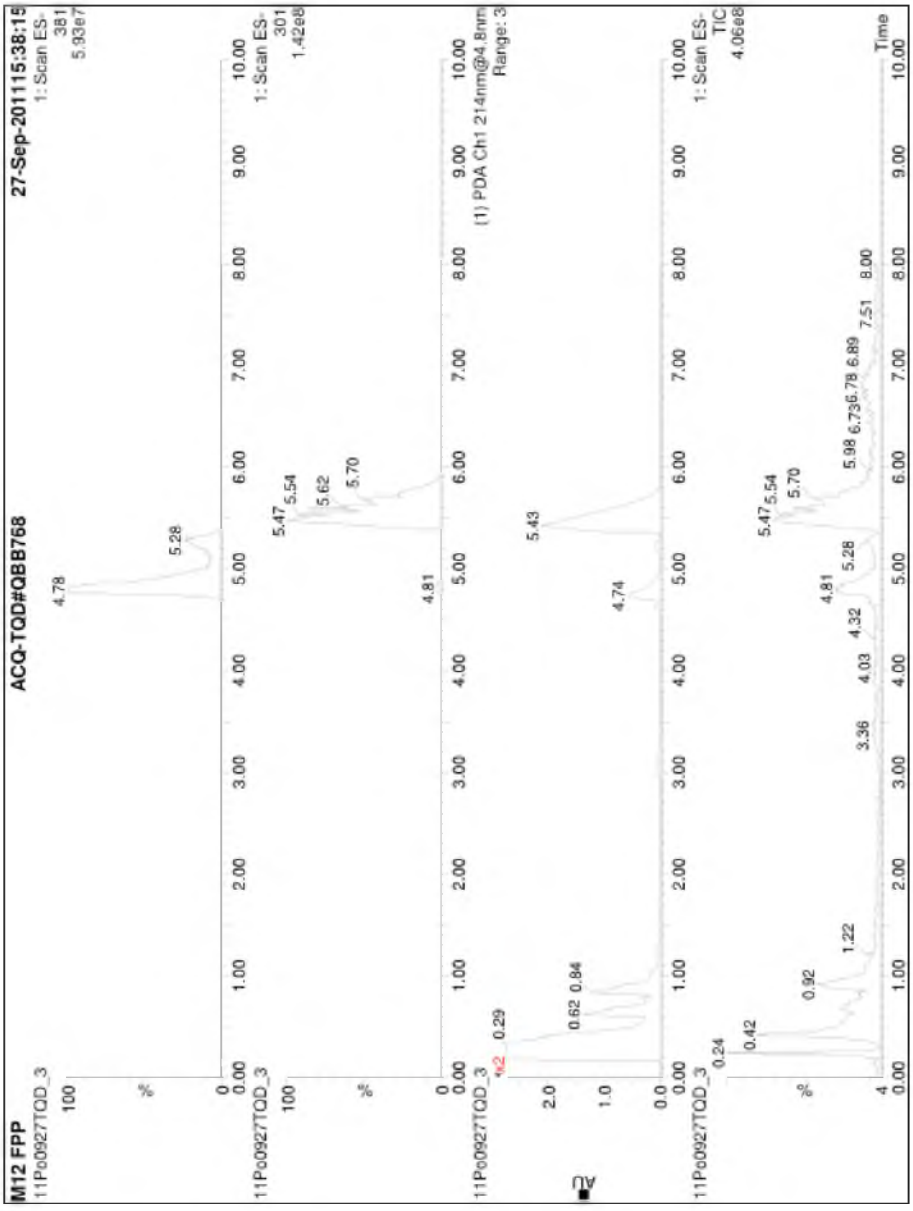


MS/MS Fragmentation of Fosfomycin Standard

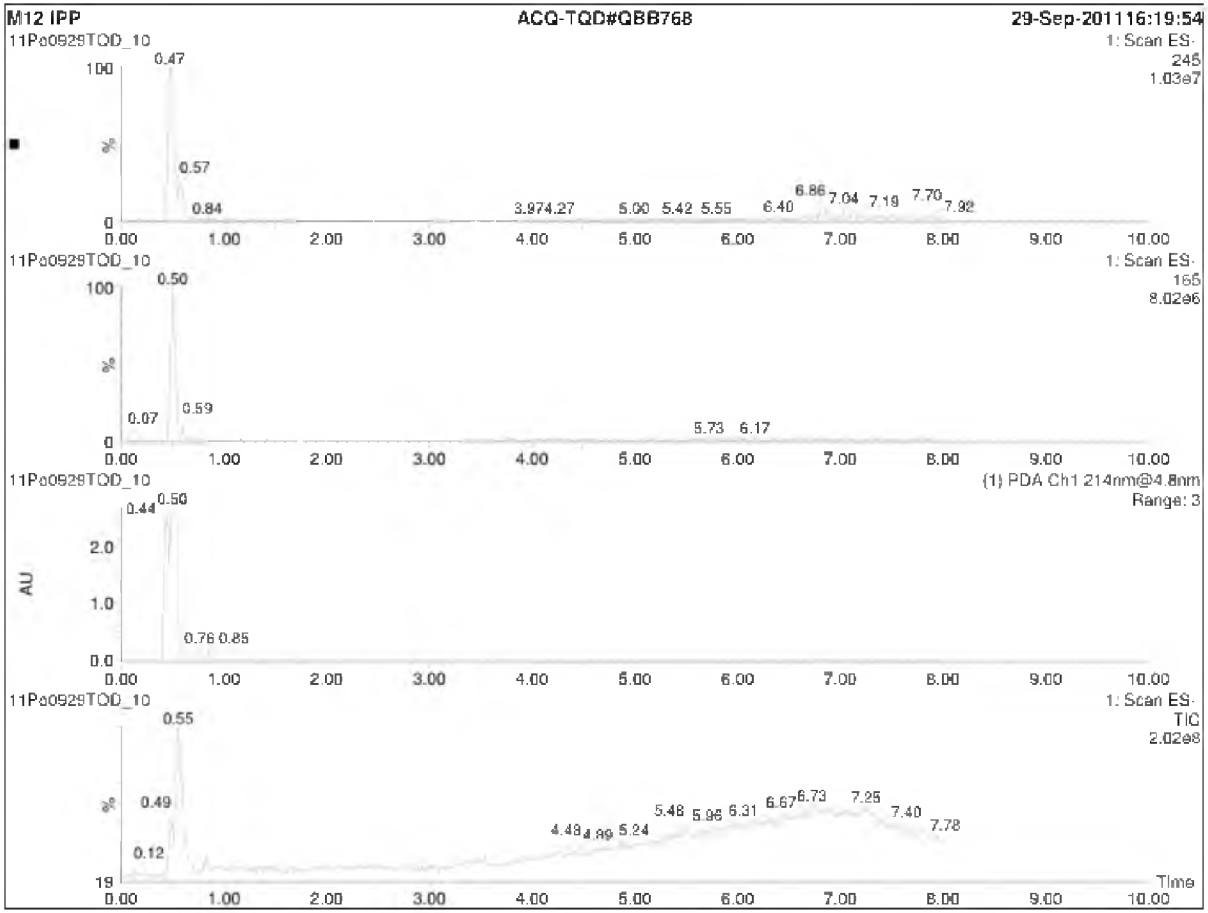
O
OH
P
+
O
120.8
115.1
m/z
110
120
130
140
150



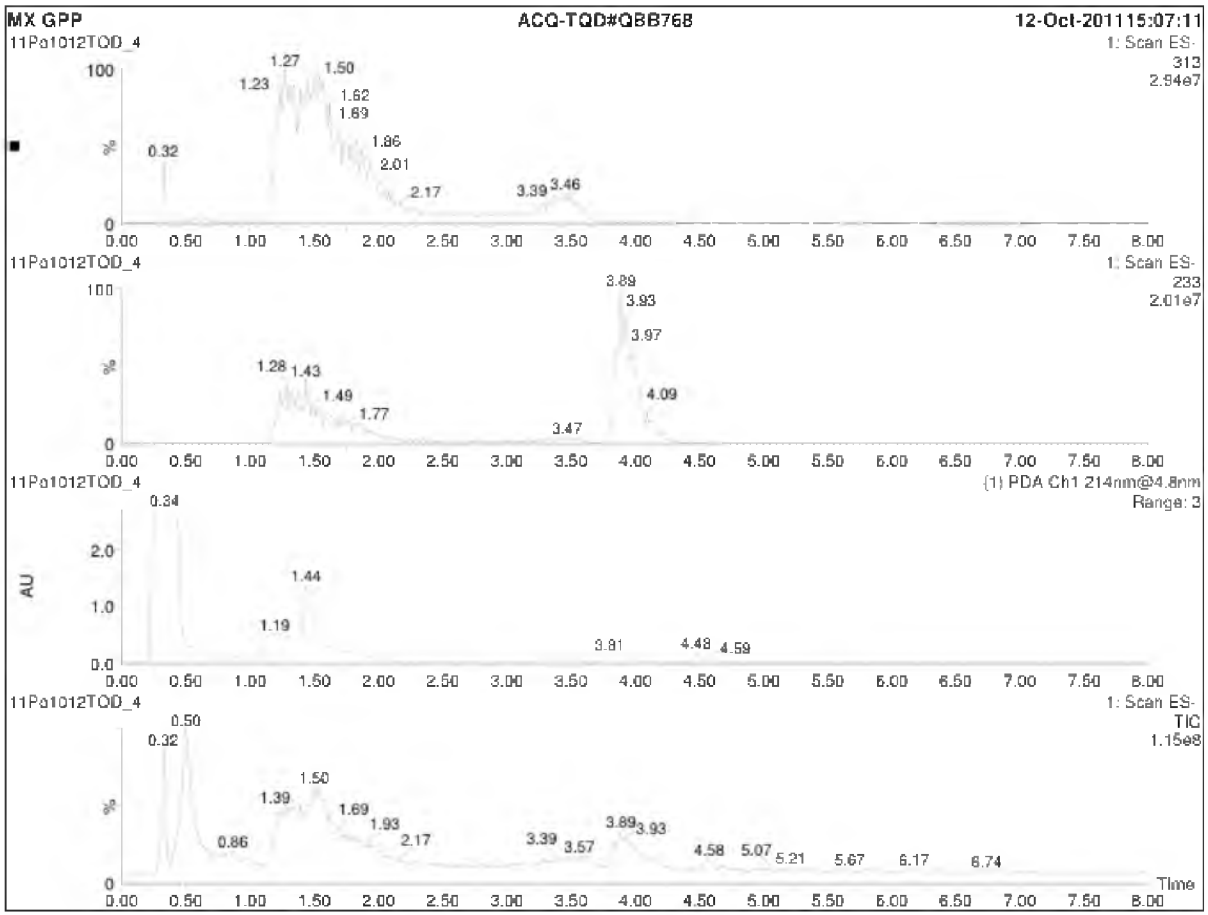
UPLC-MS of Mutant 12 using GP



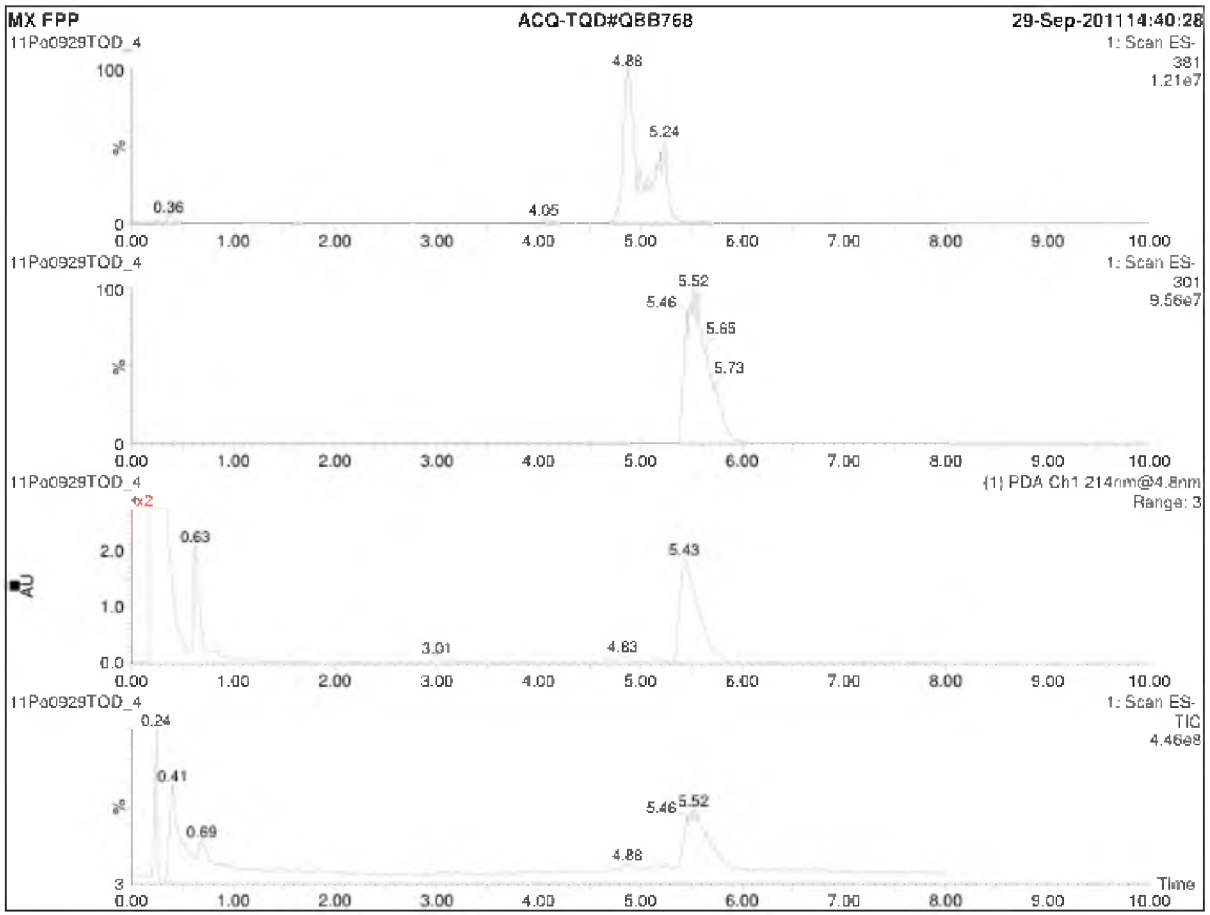
UPLC-MS of Mutant 12 using FP



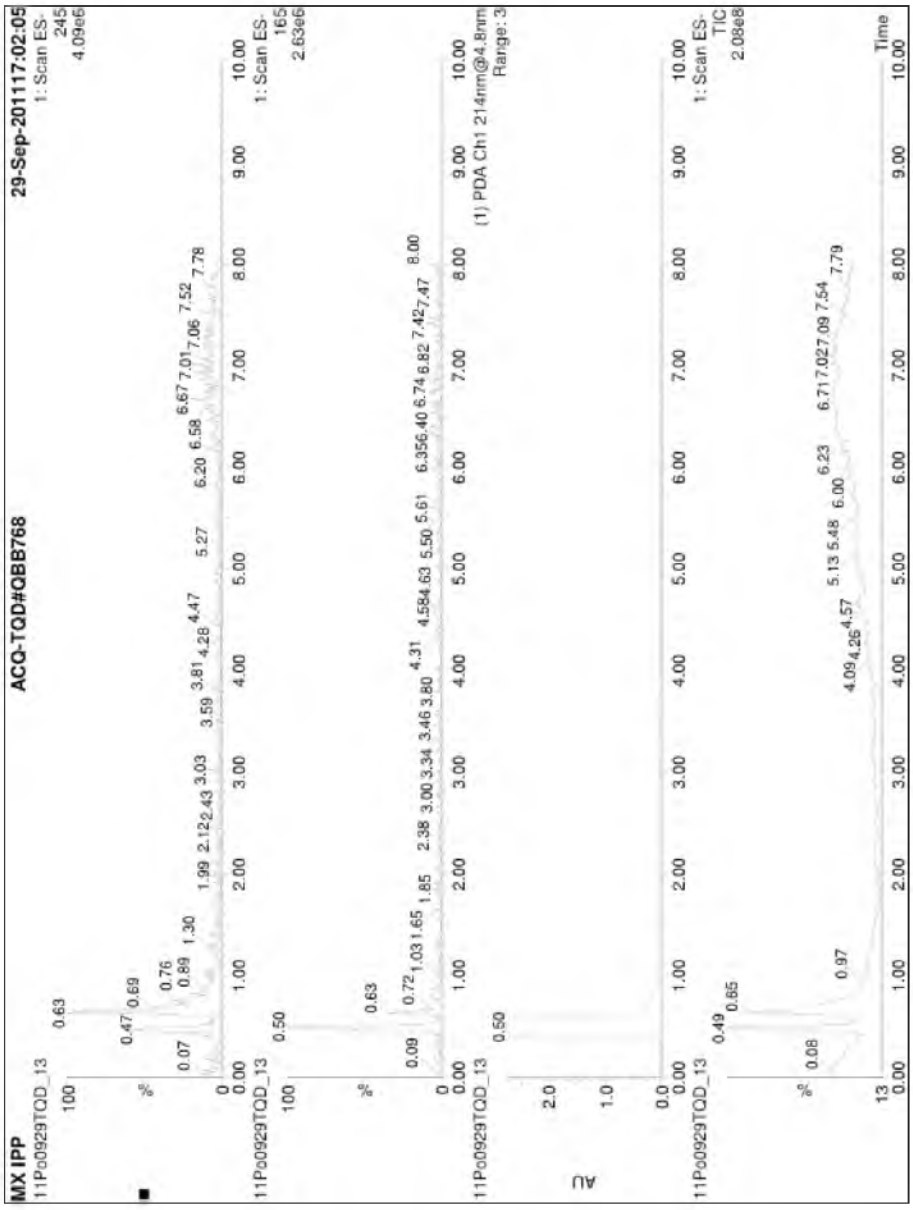
UPLC-MS of Mutant 12 using IP



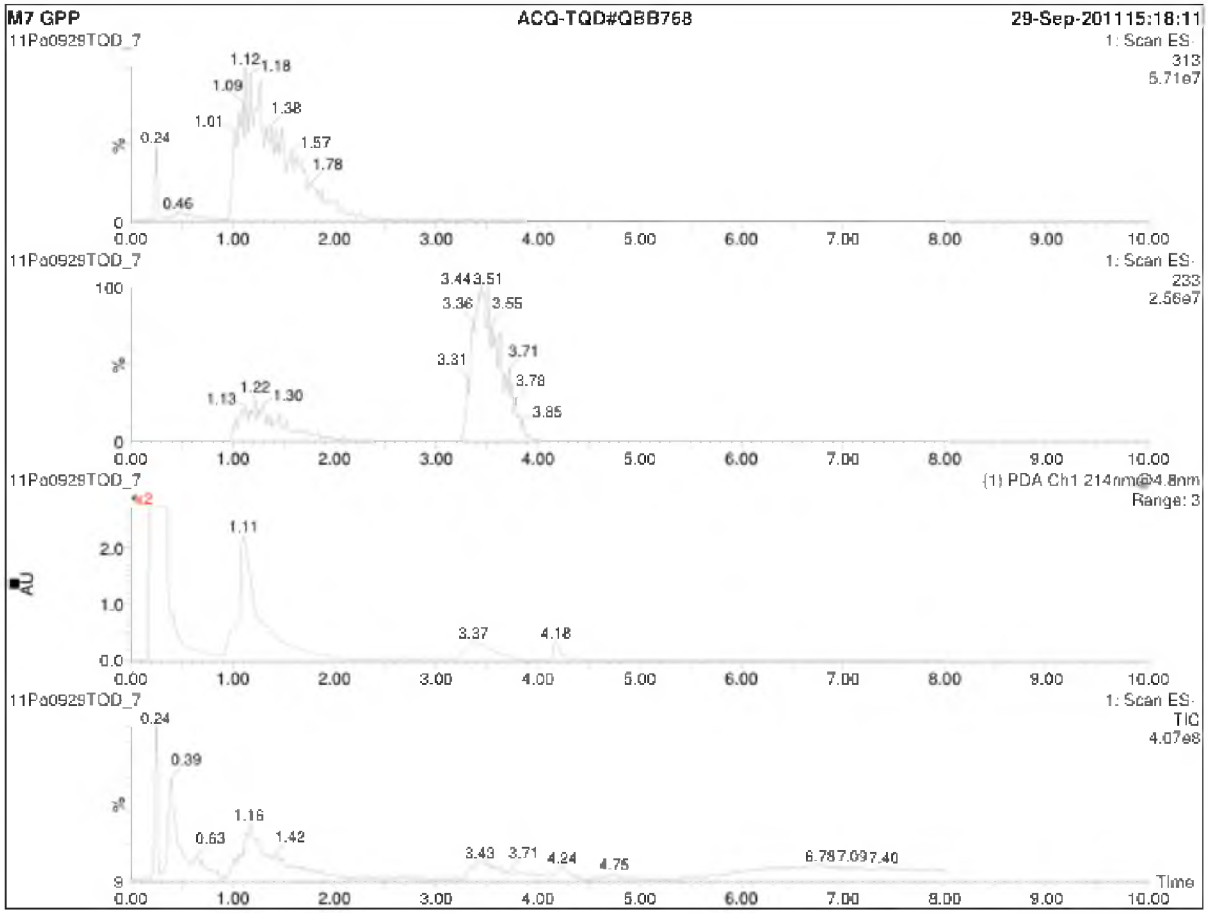
UPLC-MS of Mutant X using GP



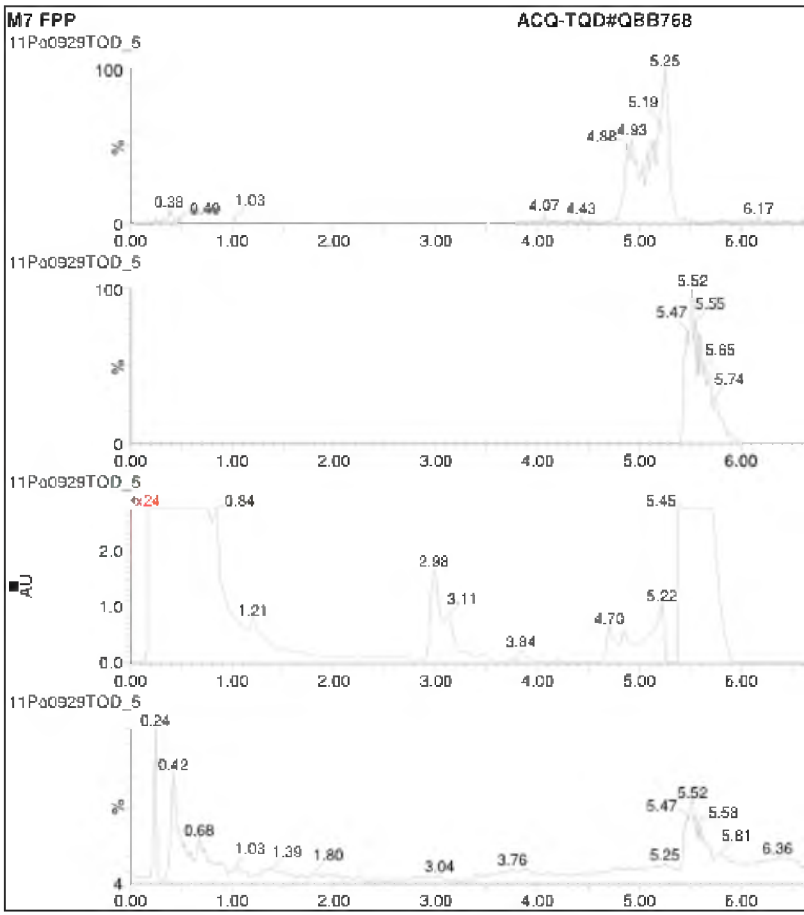
UPLC-MS of Mutant X using FP



UPLC-MS of Mutant X using IP

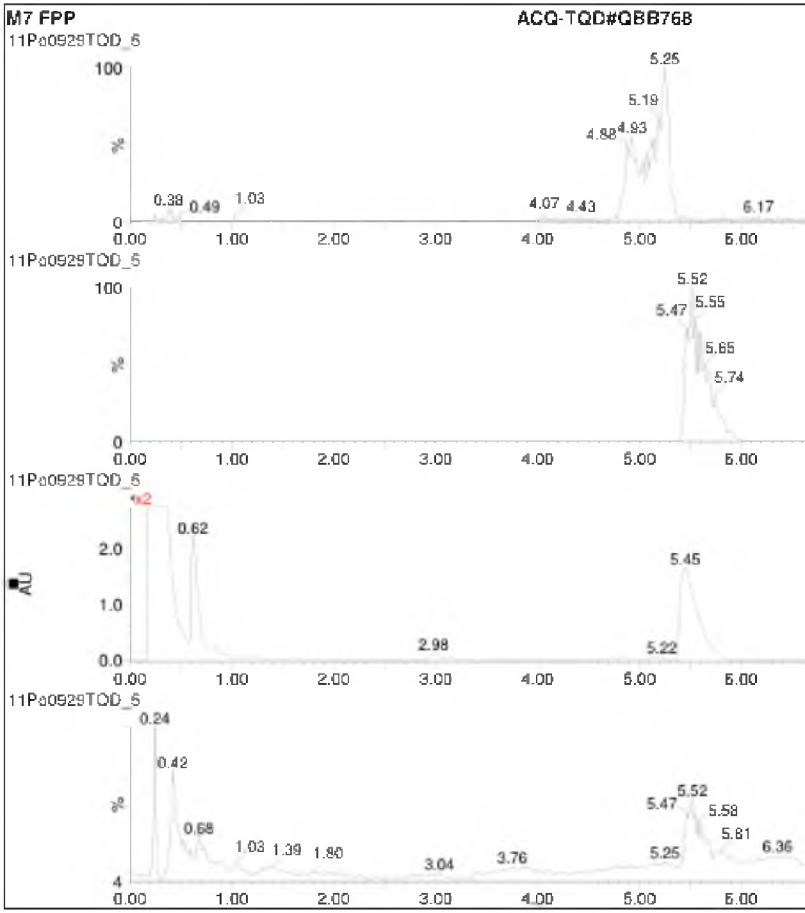


UPLC-MS of Mutant 7 using GP



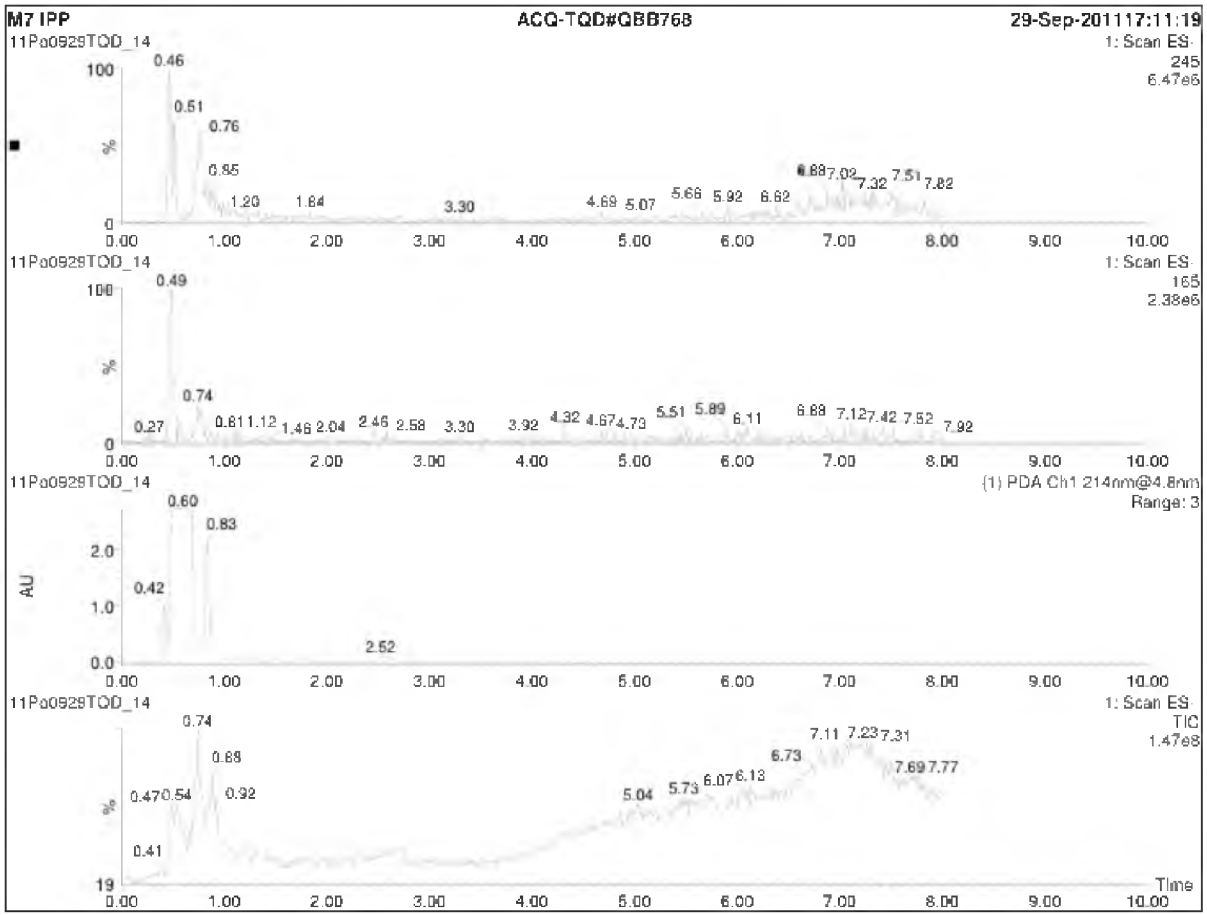
UPLC-MS of Mutant 7 using FP (x 24)

29-Sep-201114:51:35
1: Scan ES-
381
8.53e6
7.00
8.00
9.00
10.00
1: Scan ES-
301
1.08e8
7.00
8.00
9.00
10.00
(1) PDA Ch1 214nm@4.8nm
Range: 3
7.00
8.00
9.00
10.00
1: Scan ES-
TIC
4.44e8
7.00
Time
7.00
8.00
9.00
10.00

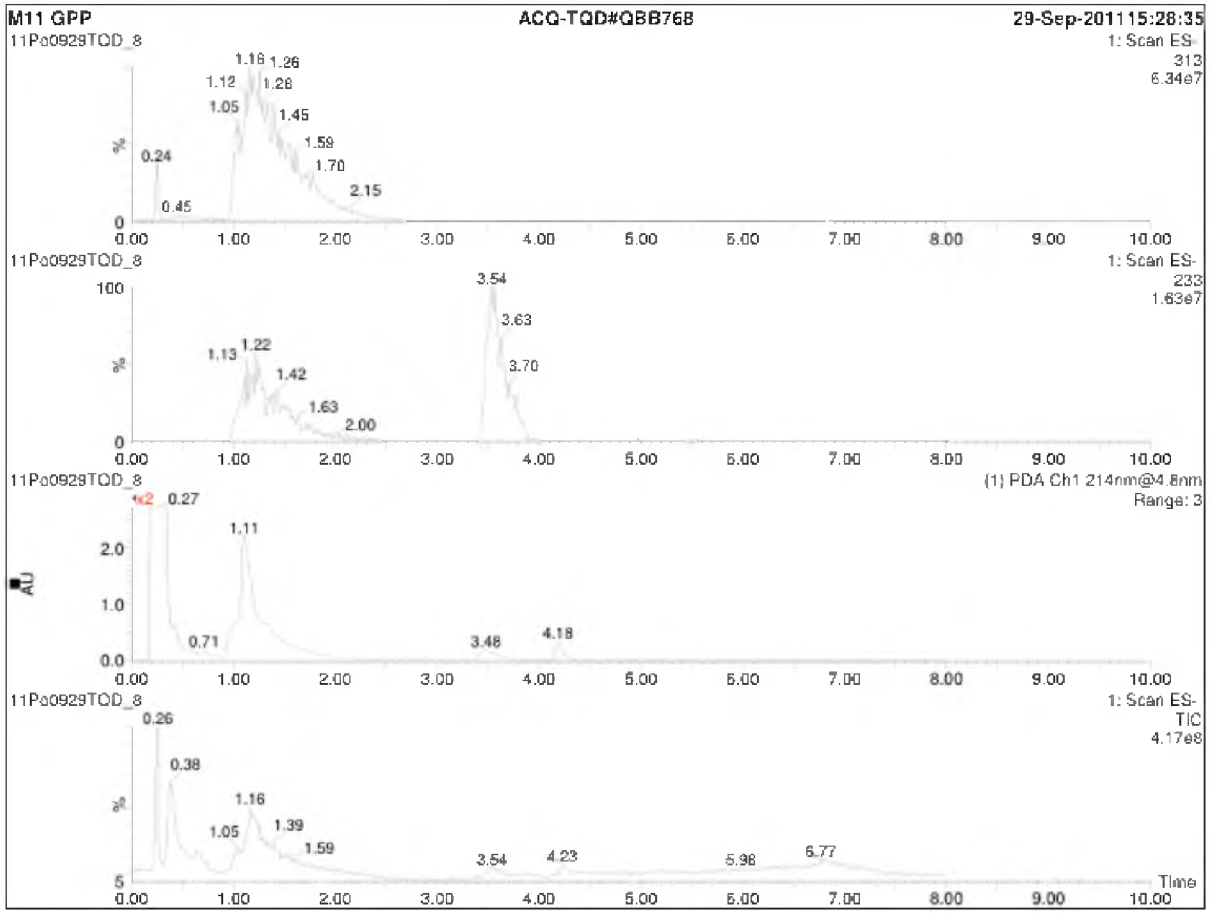


UPLC-MS of Mutant 7 using FP (x2)

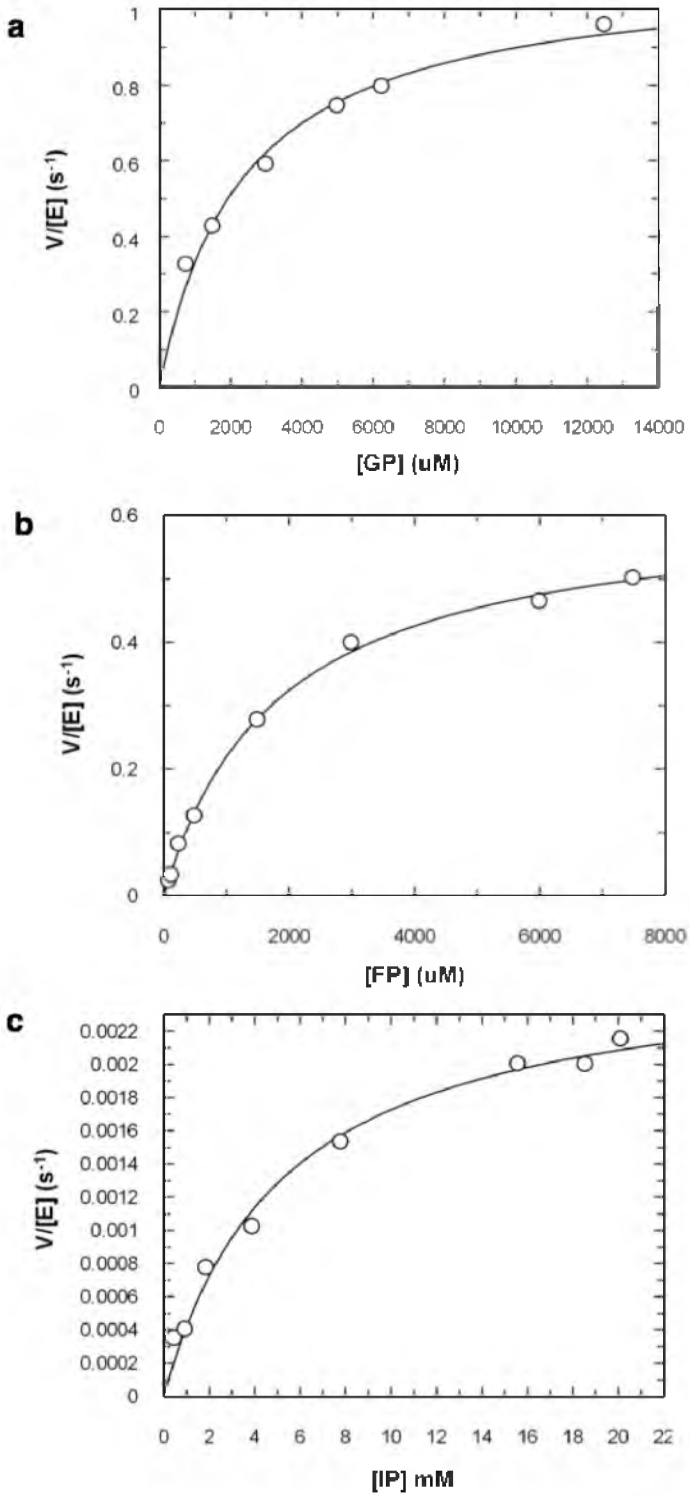
29-Sep-201114:51:35
1: Scan ES-
381
8.53e6
7.00
8.00
9.00
10.00
1: Scan ES-
301
1.08e8
7.00
8.00
9.00
10.00
(1) PDA Ch1 214nm@4.8nm
Range: 3
7.00
8.00
9.00
10.00
1: Scan ES-
TIC
4.44e8
7.00
Time
7.00
8.00
9.00
10.00



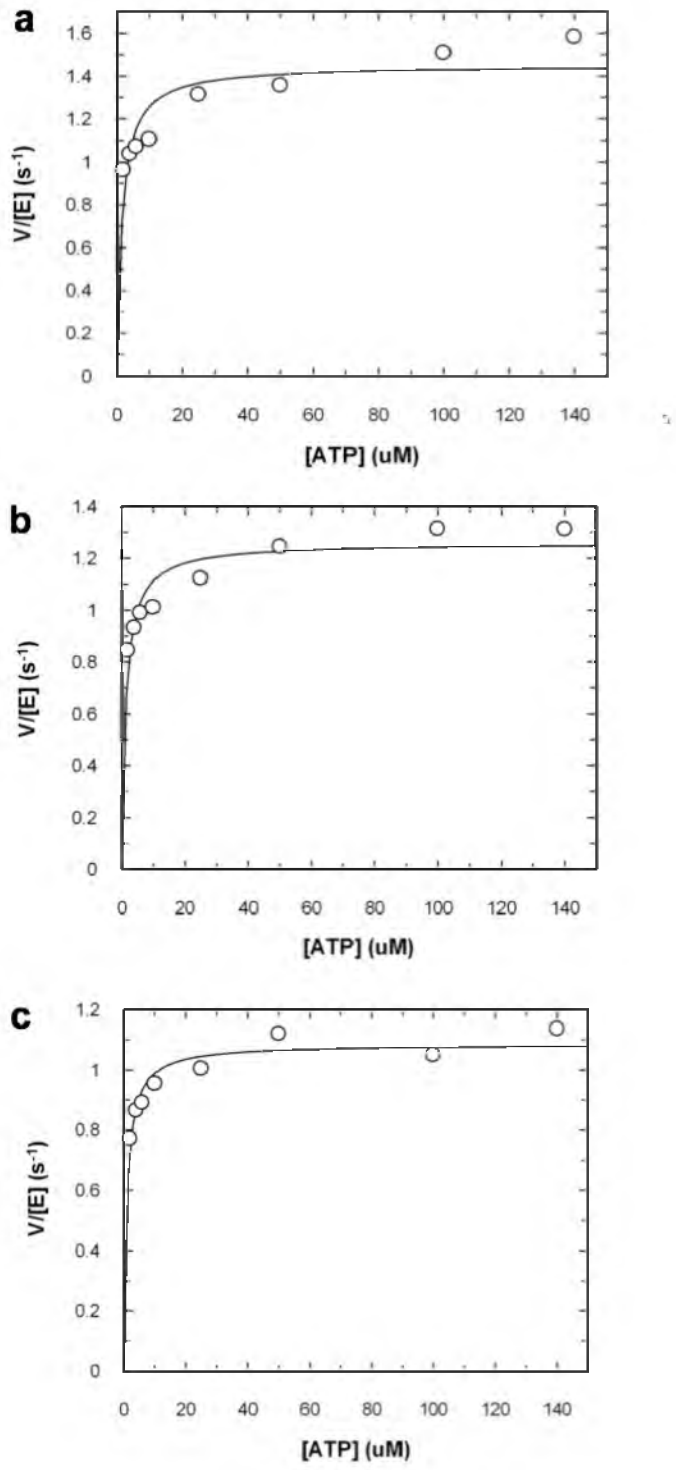
UPLC-MS of Mutant 7 using IP



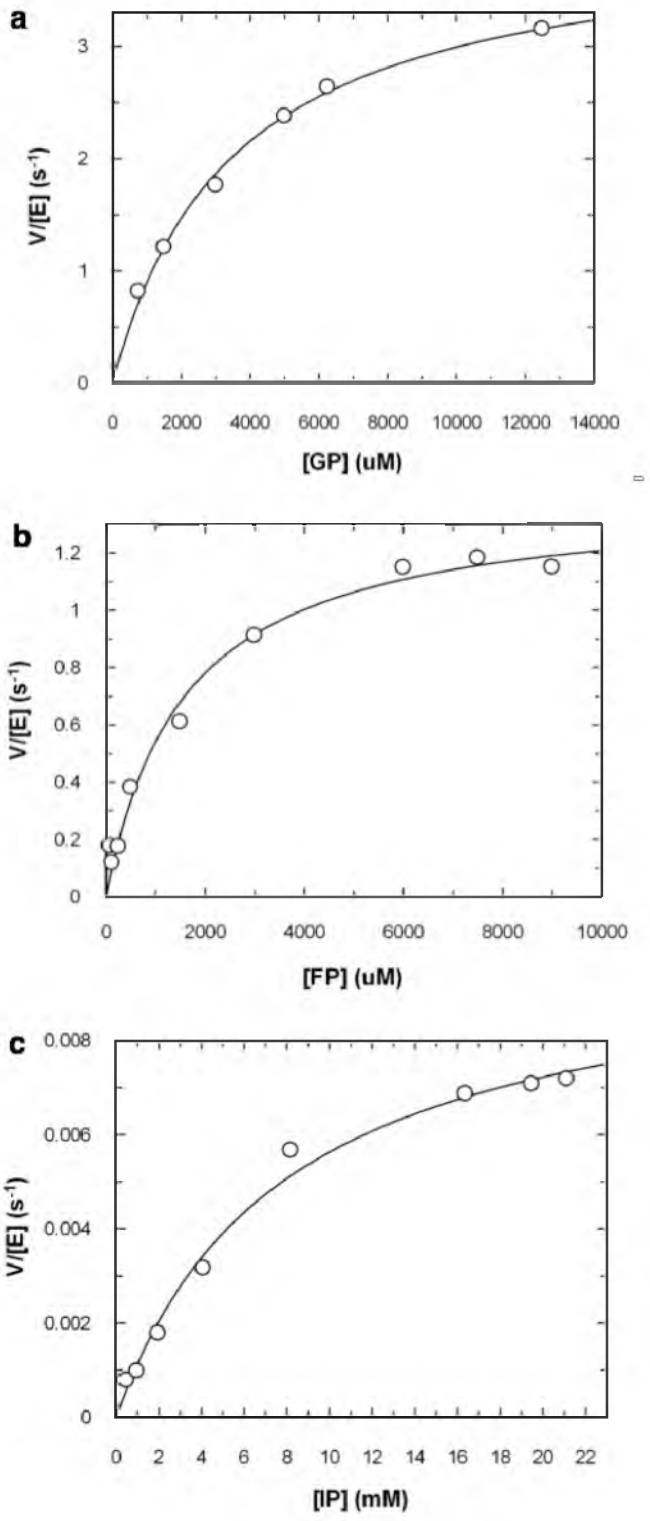
UPLC-MS of Mutant 11 using GP



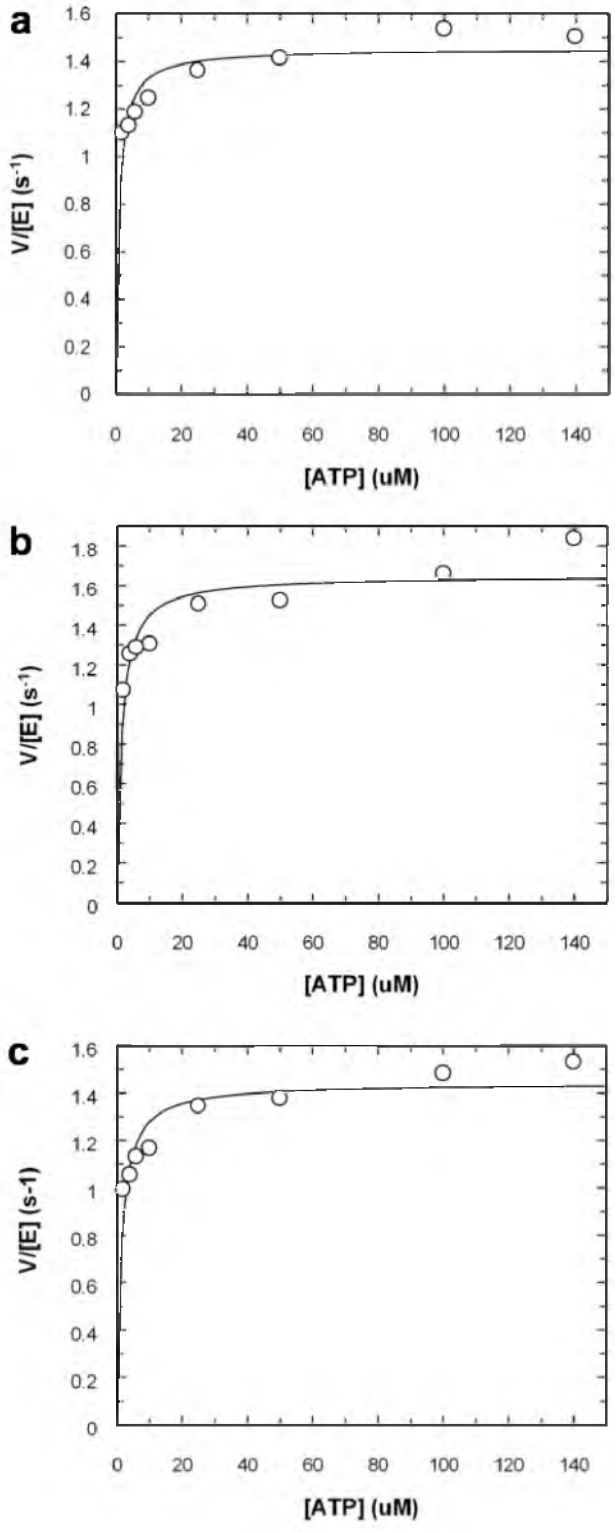
Michaelis-Menten curves for Mutant 12 using GP (a), FP (b) and IP (c) as substrates at 250 μM ATP.



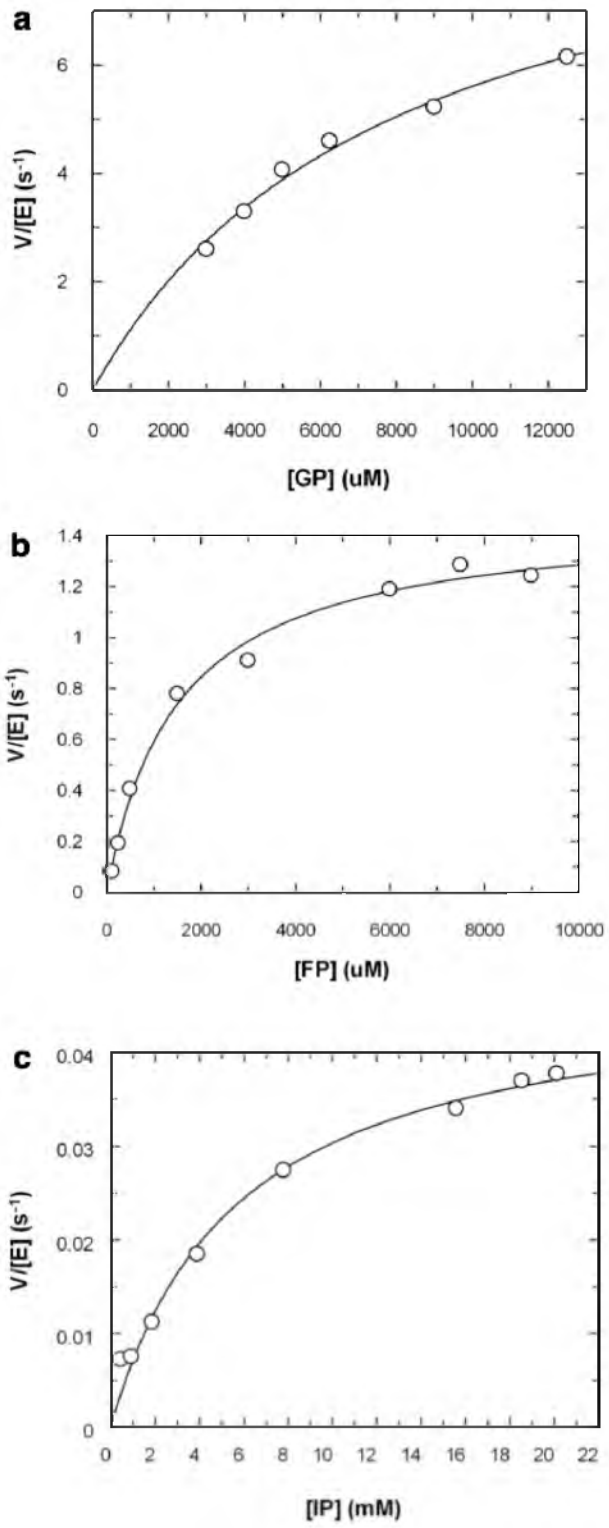
Michaelis-Menten curves for the use of ATP by Mutant 12 using 10 mM GP (a), 7 mM FP (b), and 20 mM IP (c) as cosubstrates.



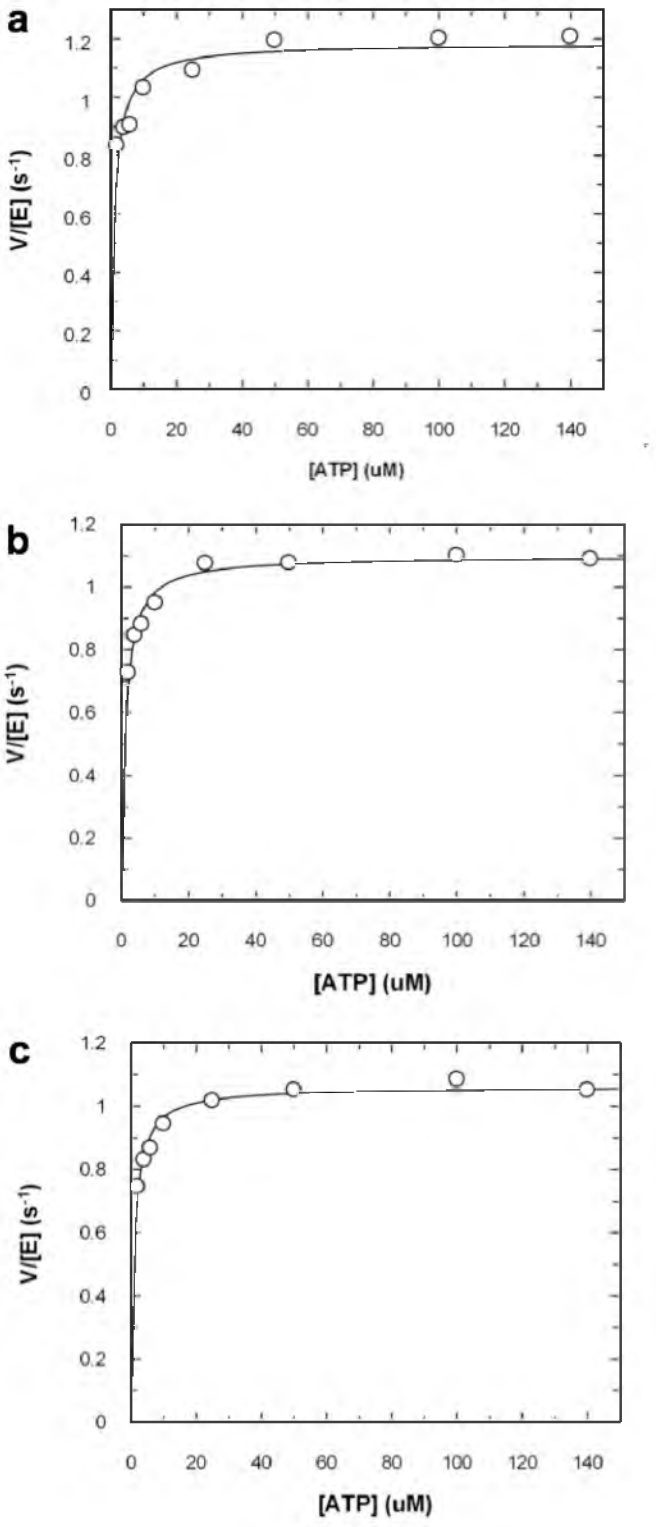
Michaelis-Menten curves for Mutant X using GP (a), FP (b), and IP (c) as substrates at 250 μM ATP.



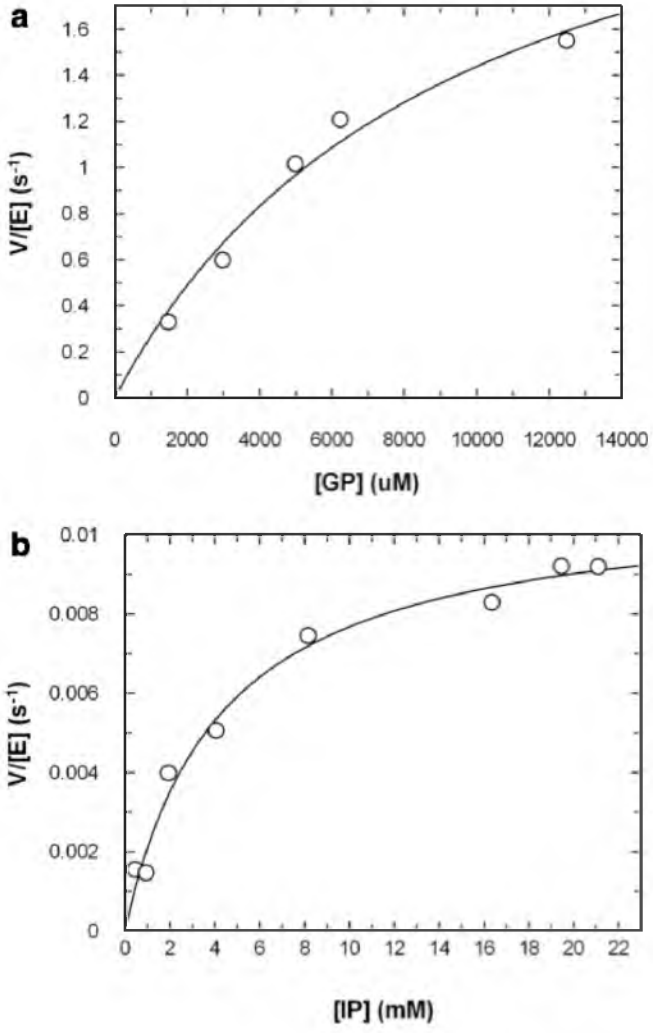
Michaelis-Menten curves for the use of ATP by Mutant X using 10 mM GP (a), 7 mM FP (b), and 20 mM IP (c) as cosubstrates.



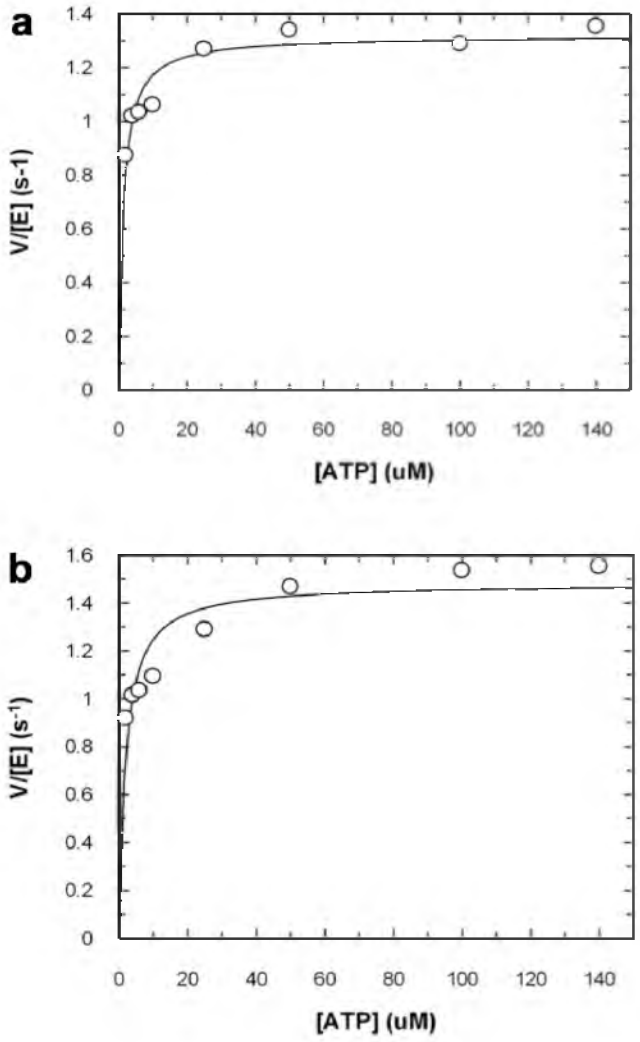
Michaelis-Menten curves for Mutant 7 using GP (a), FP (b), and IP (c) as substrates at 250 μM ATP.



Michaelis-Menten curves for the use of ATP by Mutant 7 using 10 mM GP (a), 7 mM FP (b), and 20 mM IP (c) as cosubstrates.



Michaelis-Menten curves for Mutant 11 using GP (a) and IP (b) as substrates at 250 μM ATP.



Michaelis-Menten curves for the use of ATP by Mutant 11 using 10 mM GP (a) and 20 mM IP (b) as cosubstrates.